An Investigation of *Query-by-Drawing* Image Search on Mobile Devices

Min Zhang, BSc., MEng.

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

March 2016

Abstract

The rapid growth of touchscreen mobile devices has opened up many new opportunities for exploring Content-based Image Retrieval (CBIR) techniques, e.g., it has enabled the development of intuitive and natural user interfaces to facilitate the Query-by-Drawing (QbD) image search paradigm. Although ubiquitous mobile computing holds the promise of drastically changing the ways users search for images from large image repositories, there is yet no comprehensive and systematic study that examines the factors that influence the usability, user behaviours and task performance of a QbD image search mobile application (App). No guidelines and principles exist for designing fundamental user interface components such as the colour picker and little is known how its different designs affect system usability and task performance. Little research exists that examines how different user interface designs affect the ways users draw the queries and how users go about the search process for different types of tasks. The questions of whether or not it is possible or how well users can perform QbD image search by drawing from memory of previously-seen images are unexplored. An understanding of these questions is very important and useful in the development of effective and user-friendly systems for a QbD image search App on touchscreen mobile devices.

This PhD project attempts to answer these questions and more by examining the factors that influence the usability and task performance of a QbD painting search mobile application. Starting with a comprehensive literature review and current mobile App review of various related fields, we first designed and implemented a Client-Server painting search mobile App based on an existing CBIR algorithm as the research platform for collecting empirical data; we then conducted a focus group study from whose findings we re-designed the user interfaces of our QbD App. An online survey about art preference was carried out, and we designed four comprehensive user studies and recruited a total of 123 participants to take part in the experiments. Both qualitative and quantitative measures are collected and analysed to discover the various factors that influence the usability and task performance of the interface designs of a QbD image search mobile App and the memory drawing over time. Finally, we made

recommendations and suggestions on the design and implementation of various interface components of the QbD image search mobile App based on our findings.

This thesis presents the following contributions summarised as: 1) We build a flexible platform that can be used for research in drawing-related fields, such as a new QbD technique or interface tests and psychological study. 2) We present a comprehensive and systematic review of the methods and techniques related to the investigation of QbD image search mobile App. 3) We propose a new way of categorising colour picker: '1D-1D-1D', '1D-2D', and '3D' colour picker. 4) The rationales of choosing stimuli are proposed and a real-world painting database is built. 5) We also develop a variety of novel methodologies for experimental design, data collection and data analysis, and we formulated a new protocol for assessing drawing accuracy and search result. 6) We propose some colour picker design guidelines through a series of experiments and the analysis of comprehensive experimental data. 7) Although indeed memory decays over time, we found the participants are able to draw from the memory of a painting with simple compositional structure (6-7 colour blobs), even for the paintings viewed a month ago. And finally, 8) the experiments also provide valuable insights into how general users draw and modify a query, and judge the result relevance for different tasks on mobile phones, as well as search pattern and memorisation strategy, which also extend a scientific understanding of using current Query-by-Drawing techniques for real world image search.

.

List of Publications

1. **Zhang, M.**, Qiu, G., Alechina, N. and Atkinson, S., (2015). A Preliminary Examination of the User Behavior in Query-by-Drawing Portrait Painting Search on Mobile Devices. In 13th *International Conference on Advances in Mobile Computing and Multimedia (MoMM)*, Brussels, December 2015. [Chapter 6]

2. **Zhang, M.**, Qiu, G., Alechina, N. and Atkinson, S., (2015). A Comparison of Five HSV Color Selection Interfaces for Mobile Painting Search. In 15th *International Conference on Human-Computer Interaction (INTERACT)*, Bamberg, September 2015. [Chapter 7]

3. **Zhang, M.**, Atikinson, S., Qiu, G. and Alechina, N., (2014). Can People Finger-draw Color-sketches from Memory for Painting Search on Mobile Phone? In *Proc. of 12th International Conference on Mobile Computing and Multimedia (MOMM)*, 115-118, Kaohsiung, December 2014. [Chapter 9]

4. **Zhang, M.**, Qiu, G., Alechina, N. and Atkinson, S., (2014). User Study on Color Palettes for Drawing Query on Touchscreen Phone. In *International Conference on Multimedia and Human-Computer Interaction (MHCI)*, Prague, August 2014. [Chapter 5]

5. **Zhang, M.**, Atkinson, S., Alechina N. and Qiu G., (2013). 'A is for Art' - My drawings, Your paintings. In *workshop on Multimedia for Culture Heritage (MM4CH) at 17th International Conference on Image Analysis and Processing (ICIAP)*, Naples, September 2013. [Chapter 4]

6. **Zhang, M.**, (2013). 'A is for Art' - Draw Paintings in You Mind. Presented at *Electronic Visualisation and the Arts (EVA)*, London, July 2013.

Acknowledgements

I have learnt so much new stuff during my life at the University of Nottingham, and many people have supported and helped me along the way of my PhD. This thesis would not be possible without them.

First, I would like to sincerely thank my PhD supervisors, Dr. Guoping Qiu, Dr. Sarah Atkinson, and Dr. Natasha Alechina for their wonderful supervision and guidance. They are my mentors and advisors, as well as my friends. This PhD work is shaped by frequent discussions with Dr. Qiu from the very beginning, and he has advised me how to ask right research questions and to look at a big picture, and he also gave me some suggestions on the technical issues. Dr. Atkinson and Dr. Alechina have suggested some insights on the experimental design from their diverse background. More importantly, they encouraged me to be positive when came across hard problems and without their suggestions and encouragement, this thesis could not be completed.

I want to thank Dr. Michael Brown and Dr. Max Wilson for their valuable advices in HCI, Dr. Dominic Price for offering me free Server space, David Morton for organizing mobile devices lending service, and all anonymous reviewers of my submitted paper for insightful directions. I also would like to thank all evaluators at the early development stage of my App and experiment testers, as well as all participants who have attended the experiments.

Furthermore, I am also grateful to have all members in Horizon and IMA research group, a special thanks to Dr. Steve Benford, Dr. Mike Sharples, Dr. Claire O'Malley, Dr. Sarah Sharples, Emma Juggles, and Nadine Holmes. Thanks to all my colleagues in Visual Information Processing group: Dr. Hao Fu, Dr. Mercedes Torres, Dr. Jie Shu, Dr. Salvador Garcia Bernal, Dr. Qian Zhang, Bozhi Liu, Chao Zhang, Zhan Xu, Chi Zheng, Zhuo Lei, Bolei Xu, Jingxin Liu, Yang Hao and Dr. Chuan Zhou. Thanks to all the friends who have made this truly unforgettable life experience in Nottingham: Sameh Zakhary, Venkatasuresh Bonthala, Elissa Madi, Xiaomeng Wang, Bo Wang, Wenwen Bao, Jianhua Shao, Yin He, Xia Li, Ning Xue, Yao Li, Yiming Quan, Liming Xu, Shuli Liu, Mengqi Shen, Lanyun Zhang, Boying Li, Mengdi Li, Xiaojun

Li, Xinyu Fu, Edu, Marian, Vasu, and Elisa Tonello. Thanks to all friends who I have met on the conference and summer school, for sharing their experience of both research and life which broaden my vision and possibility.

I would also like to thank all of my friends and family in China, especially my mother and farther, my sister and my boyfriend who have supplied me unconditional support and boundless love.

Finally, my biggest thanks to China Scholarship Committee, Horizon CDT and the University of Nottingham for offering a scholarship to support my PhD work and life, without these this thesis is not possible.

Contents

Abstract i

List of Publications iii

Acknowledgements iv

Contents vi

List of Figures xv

List of Tables xxiv

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Motivation and Objectives . . . 4
- 1.3 Research Questions . . . 6
- 1.4 Contributions . . . 6
- 1.5 Organization of the Thesis . . . 9

2 Review of Related Work and Methods 13
- 2.1 Introduction . . . 13
- 2.2 Colour Selection and Colour Picker . . . 14
  - 2.2.1 Colour Model . . . 15
  - 2.2.2 Colour Picker: User Interface for Colour Selection . . . 17
    - 2.2.2.1 Representation of Colour Picker . . . 17
      - Colour Swatch . . . 17
      - Slider Bar . . . 18
      - Colour Ring . . . 19
      - Colour Wheel . . . 21
      - Colour Square . . . 22

2.2.2.2 Colour Picker of Combined Shapes 23
Bar-Wheel 24
Bar-Square 24
Ring-Square/Triangle 26
More than Two Shapes 27
2.2.2.3 Colour picker Categories 29
3D Colour Picker 31
2.2.3 Existing Colour Pickers 32
2.2.3.1 Colour Pickers used in CBIR 32
Colour Picker used in Query-by-Colour System . 32
Colour Pickers used in Query-by-Drawing System 35
2.2.3.2 Colour Pickers used in Current Mobile Applications 38
2.2.4 A Survey on Mobile Phone Use and Drawing Frequency 41
2.3 Related Work of Colour Selection Study 42
2.3.1 Colour Matching Experiment 42
2.3.2 Visual Design of Colour Picker 44
2.3.3 Colour Selection Study in CBIR 47
2.3.4 Colour Selection Strategy 49
2.4 Drawing-Related work 51
2.4.1 Psychological Studies on Drawing 52
2.4.2 Drawing as Query Input 53
2.4.2.1 Drawing Query on Computer 53
2.4.2.2 Drawing Query on Mobile Devices 57
2.4.3 Drawing Behaviour Study 57
2.4.4 Current Mobile Drawing Applications 59
2.5 Drawing from Memory 61
2.5.1 Colour Memory Study 61
2.5.2 Memory Drawing Study 62
2.6 Summary 64

3 Introducing our *Query-by-Drawing* Mobile Application 66
3.1 Introduction 66
3.2 Overview of System 67
3.2.1 Client-Server Architecture 67
3.2.2 Image Matching Algorithm 68
3.2.3 Database 70
3.3 User Interfaces of QbD System 71
3.3.1 Focus Group Study on Early-stage of Prototype UI . . . 71
3.3.1.1 Early Prototype User Interfaces 72
3.3.1.2 Participants 74

3.3.1.3 Procedure . . . 74
3.3.1.4 Results and Discussion . . . 75
3.3.2 Revised User Interface of Basic Tools . . . 78
3.3.3 Revised Colour Picker Interface . . . 79
3.3.4 Result display Interface . . . 80
3.4 Discussion . . . 82
3.5 Summary . . . 82

4 Research Design and Methodology 83
4.1 Introduction . . . 83
4.2 Study Overview . . . 84
4.3 Experimental Planning . . . 86
4.3.1 Determining Stimuli . . . 86
4.3.1.1 An Survey on Painting Preference . . . 87
4.3.1.2 Stimulus Choosing Rationales . . . 88
4.3.2 Participant Recruitment . . . 89
4.3.3 Ethical Approval . . . 90
4.4 Methods for Experimental Execution . . . 90
4.4.1 Experimental Setting . . . 90
4.4.2 Experimental Task Type . . . 91
4.4.3 Experimental Procedure . . . 93
4.5 Methods for Data Collection and Data Analysis . . . 94
4.5.1 Measurements Definition . . . 94
4.5.2 Independent variables . . . 96
4.5.3 Data Collection Methods . . . 97
4.5.3.1 Think-Aloud Protocol . . . 97
4.5.3.2 Questionnaires . . . 97
Pre-Questionnaire . . . 97
Post-Questionnaire . . . 97
4.5.3.3 Interview . . . 100
4.5.3.4 Search Video Logs . . . 100
4.5.3.5 Drawings Ratings . . . 101
Human Accuracy Rating . . . 101
Coding Portrait Drawing . . . 102
Coding Landscape Drawing . . . 102
4.5.4 Data Analysis Methods . . . 103
4.6 Justifications of Research Design . . . 104
4.7 Limitations . . . 105
4.8 Summary . . . 106

5 Experiment 1 - Discrete and Continuous Colour Picker 108
5.1 Introduction . . . 108

5.2 Aims of Experiment . . . 109
5.3 Colour Picker Interfaces . . . 110
5.3.1 Database and Results Displa . . . 110
5.3.2 Discrete Colour Pickers . . . 111
5.3.3 Continuous Colour Pickers . . . 113
5.4 Experimental Design and Procedure . . . 115
5.4.1 Stimuli . . . 115
5.4.2 Participants . . . 116
5.4.3 Procedure . . . 116
5.5 Results . . . 118
5.5.1 Statistical Analysis of Copying-tasks . . . 118
5.5.1.1 Time of Drawing a Query . . . 119
5.5.1.2 Number of Colours Depicted in a Query . . . 120
5.5.1.3 Number of Objects Depicted in a Query . . . 121
5.5.1.4 Search Performance . . . 122
5.5.1.5 Human Ratings on Colour Accuracy . . . 124
Query's Colour Accuracy . . . 124
Foreground Colour Accuracy . . . 125
Background Colour Accuracy . . . 125
5.5.1.6 Participants' Usability Choices . . . 126
5.5.1.7 Preference for Colour Picker . . . 128
5.5.2 Drawing Pattern . . . 130
5.5.2.1 First Interaction with QbD System . . . 130
5.5.2.2 Drawing Sequence . . . 131
5.5.2.3 Uses of Tools . . . 133
5.5.3 Query Refinement . . . 133
5.5.4 Search Strategies . . . 135
5.5.5 Opinions and Suggestions on Overall System . . . 136
5.6 Discussion . . . 137
5.6.1 Effects of Colour Picker Type . . . 137
5.6.2 Effects of System Design . . . 140
5.6.3 Effects of User Characteristics . . . 142
5.6.4 Drawing Process . . . 143
5.7 Summary . . . 144

6 Experiment 2 - Task Type, Painting Type and Brush Type 146
6.1 Introduction . . . 146
6.2 Aims of Experiment . . . 148
6.3 Colour Picker Interfaces . . . 149
6.3.1 Database and Result Displaying . . . 150
6.3.2 Discrete Colour Pickers . . . 150

6.4 Textural Brush System Design . . . 152
6.4.1 General User Interfaces . . . 153
6.4.2 Texture Palette of '*Portrait*' . . . 154
6.4.3 Texture Palette of '*Landscape*' . . . 154
6.5 Experimental Design and Procedure . . . 154
6.5.1 Stimuli . . . 156
6.5.2 Participants . . . 156
6.5.3 Procedure . . . 156
6.6 Results . . . 159
6.6.1 Effects of Colour Picker Type and Combinations . . . 160
6.6.1.1 Statistical Analysis of Copying-tasks on Colour Systems . . . 160
6.6.1.2 Statistical Analysis of All Tasks on Colour Systems . . . 162
6.6.2 Effects of Task Type . . . 164
6.6.3 Effects of Painting Type - Portrait vs. Landscape . . . 166
6.6.4 Effects of Brush Type - Plain colour vs. Texture . . . 166
6.6.5 Real-life Use Scenario - T2 tasks and T3 tasks . . . 168
6.6.5.1 Ethnic Group Effects . . . 168
6.6.5.2 Statement Agreement Ratings . . . 171
6.6.5.3 System Choices and Task Type . . . 174
T2-Portrait tasks and T3 tasks . . . 175
Seasonal Landscape Painting Queries . . . 176
Participants' System Choosing Criteria . . . 177
6.6.5.4 Use of System Texture . . . 178
6.6.6 Drawing Pattern for Target-unspecific Tasks . . . 179
6.6.6.1 Aspect Ratio . . . 179
6.6.6.2 Perspective . . . 180
6.6.6.3 Background . . . 181
Background-First . . . 181
Mixed Background . . . 181
Complex Background Colour . . . 182
No *Background* . . . 182
6.6.6.4 Inclusion of Objects . . . 183
Drawing from Description (T2 tasks) . . . 183
Drawing after 30-second Viewing (T3 tasks) . . . 185
6.6.6.5 Cultural Background and Experience . . . 185
T2-Portrait tasks . . . 185
T2-Landscape tasks . . . 186
T3 tasks - Drawing after 30-second Viewing . . . 187
6.6.7 Participants' Judgements on Relevance . . . 187

6.6.7.1 Specific Objects or Colours 188
6.6.7.2 Compositional structure 190
6.7 Discussion 191
6.7.1 Verifying Findings of Experiment 1 191
6.7.2 Effects of Task Type 194
6.7.3 Effects of Painting Type 195
6.7.4 Effects of Brush Type 195
6.7.5 Participant Characteristics 197
6.7.6 Drawing Pattern 198
6.8 Summary 200

7 Experiment 3 - Visual Design Attributes of Continuous Colour Picker 202
7.1 Introduction 202
7.2 Aims of Experiment 203
7.3 Database and Results Displaying 204
7.4 Colour Picker Interfaces 204
7.4.1 H-SVsquare Colour Picker 205
7.4.2 S-HVsquare Colour Picker 206
7.4.3 V-HSsquare Colour Picker 207
7.4.4 V-HSwheel Colour Picker 208
7.4.5 Vring-HSwheel Colour Picker 208
7.5 Experimental Design and Procedure 209
7.5.1 Stimuli 209
7.5.2 Participants 210
7.5.3 Procedure 210
7.6 Results 211
7.6.1 Colour Parameter Combinations - H-SV, V-HS, S-HV . 212
7.6.1.1 Time of drawing a Query 212
7.6.1.2 Number of Colours Depicted in a Query . . . 213
7.6.1.3 Number of Objects Depicted in a Query . . . 214
7.6.1.4 Search Performance 215
7.6.1.5 Human Ratings on Colour Accuracy 216
7.6.1.6 Participants' Usability Ratings 217
7.6.1.7 Preference and Best-Worst Choices of Colour Picker 218
7.6.2 Shape of 1D Input Area - Bar vs half-Ring 219
7.6.2.1 Time of drawing a Query 219
7.6.2.2 Number of Colours Depicted in a query 220
7.6.2.3 Number of Objects Depicted in a query 221
7.6.2.4 Search Performance 222
7.6.2.5 Human Ratings of Colour Accuracy 224

7.6.2.6 Participants' Usability Ratings 225
7.6.2.7 Preference and Best-Worst Choices of Colour Picker 226
7.6.3 Shape of 2D input Area - *Square* vs *Wheel* 227
7.6.3.1 Time of drawing a Query 227
7.6.3.2 Number of Colours Depicted in a query 228
7.6.3.3 Number of Objects Depicted in a query 229
7.6.3.4 Search Performance 229
7.6.3.5 Human Ratings of Colour Accuracy 230
7.6.3.6 Participants' Usability Ratings 231
7.6.3.7 Preference of Colour Picker 232
7.7 Discussion 232
7.7.1 Colour Parameter Combination 233
7.7.2 Shape of 1D Input Area 234
7.7.3 Shape of 2D Input Area 235
7.8 Summary 236

8 Experiment 4 - A Longitudinal Memory Drawing Study 237
8.1 Introduction 237
8.2 Aims of Experiment 238
8.3 Colour Picker Interfaces 239
8.4 Database and Result Display 240
8.5 Experimental Design and Procedure 240
8.5.1 Stimuli 240
8.5.2 Participants 241
8.5.3 Procedure 242
8.6 Results 243
8.6.1 Statistical Analysis 243
8.6.1.1 Time of drawing a Query 245
8.6.1.2 Number of Colours Depicted in a query 246
8.6.1.3 Number of Objects Depicted in a query 247
8.6.1.4 Search Performance - Top51 Query 248
8.6.1.5 Participants' Similarity Ratings 250
8.6.1.6 Evaluators' Accuracy Ratings 252
Colour Accuracy of Query 252
Accuracy of Overall Query 253
8.6.2 Memorisation Strategies 255
8.6.3 Drawing Pattern 257
8.6.3.1 Decayed Memory 257
8.6.3.2 False Memory 258
8.6.3.3 Small Details 259

8.6.3.4 Experience-based Memory Drawing 260
8.7 Discussion 261
8.7.1 Effects of Delayed Time 261
8.7.2 Effects of Participant Characteristics 263
8.7.3 Effects of Painting Type 264
8.8 Summary 264

9 Conclusions **266**
9.1 Main Contributions 266
9.1.1 A QbD Prototype as Platform for Future Research . . . 267
9.1.2 Methodology 267
9.1.2.1 Measures 268
9.1.3 User Interface Design for Colour Picker 268
9.1.3.1 Colour Picker Type - Discrete vs. Continuous 268
9.1.3.2 Combinations of Discrete and Continuous Colour Picker 269
9.1.3.3 Visual Design Attributes of Continuous Colour Picker 270
Colour Parameter Combinations 270
Shape of 1D Input Area 271
Shape of 2D Input Area 271
9.1.4 Brush Type Design 271
9.1.5 Task Type and Painting type 272
9.1.6 User Characteristics 273
9.1.7 Drawing and Searching Behaviour 274
9.1.7.1 Drawing Pattern 274
9.1.7.2 Search Strategy 276
9.1.8 Viability of Drawing from Memory 276
9.1.9 Design Best Practice 278
9.2 General Limitations and Future Work 279
9.2.1 Extensions of Prototype Application 279
9.2.1.1 Image Matching Techniques 279
9.2.1.2 User Interface 279
9.2.2 Support for Database Images 280
9.2.3 Further Experiments 280
9.3 Discussion 281

A Experiment Forms **283**
A.1 Consent Form 284
A.2 T2 Task Descriptions 285
A.3 Pre-Questionnaire 286
A.4 Post-Questionnaire Q3 288

A.5 Post-Questionnaire Q5 289

B K-means Colour Values **290**

Bibliography **294**

List of Figures

2.1 Representation of a *hue* wheel, a *hue* bar, a *saturation* bar, and a *value* bar. . . . 16
2.2 Screenshot of the standard colour palette in *MS Office*^{TM} Suite. . 18
2.3 Colour swatches used in *Photoshop*^{TM}, *Retrievr* and *Search Google by Drawing*. . . . 18
2.4 RGB sliders and CMYK sliders used in *Illustrator*^{TM}. . . . 19
2.5 HSV (or HSB) sliders in *Illustrator* (left) and HSL sliders in *DrawPlus X5*. . . . 20
2.6 Colour rings with different colour models: RYB, RGB, and CMYK. . . . 20
2.7 Screenshots of three colour rings extracted from the reviewed current mobile Apps. . . . 21
2.8 Representations of a HSV cone and a *hue-saturation* colour wheel. 21
2.9 Screenshot of the colour picker used in *Apple's Mac OS X* system. 22
2.10 Screenshots of the examples of an original colour wheel (left), a colour wheel with 90 degree anticlockwise rotation (middle), and a colour wheel with 90 degree clockwise rotation (left) used in three mobile Apps. . . . 22
2.11 Representations of *green-blue, red-green*, and *red-blue* colour squares. . . . 23
2.12 Representations of *hue-lightness* (left) and *hue-value* (right) colour squares, with *saturation*=255. . . . 23
2.13 Representations of *S-HV*, *V-HS*, *S-HL*, and *L-HS* colour wheel. . 24
2.14 Picking a colour by using *V-HS* colour wheel. . . . 24
2.15 Representations of *H-SV*, *S-HV*, and *V-HS* colour square. . . . 25
2.16 Screenshots of colour pickers in *MS Paint*^{TM} (left) and *MS Word*^{TM} 2010 (right). . . . 25
2.17 Screenshots of *ColorZilla* colour picker. . . . 26
2.18 Representation of a H-SV *ring-square* (left) and H-SV *ring-triangle* (right). . . . 26
2.19 Screenshots of *Krita* colour picker (left), *My Paint Color Triangle* (middle), and *GTK+* colour picker (right). . . . 27
2.20 Screenshots of the HSV colour picker of *Colourlovers*. . . . 28

2.21 Screenshot of *ColorImpact* colour picker. 28
2.22 Screenshots of user interfaces of *CoolPicker* plugin for *3DS Max*. 29
2.23 Brewer palettes in *RColorBrewer* package of R. 30
2.24 Screenshot of 3D colour picker of *ContrastA*. 31
2.25 The colour palettes used to filter keyword-based search results on *Google Images*. 33
2.26 Screenshot of *Shutterstock Spectrum*. 33
2.27 Screenshot of *Shutterstock Palette*. 34
2.28 Screenshot of *Multicolr* from *TinEye Labs*. 34
2.29 Screenshot of *xcavator.net*. 34
2.30 Screenshots of the query canvas of *ColorMap* (left), *PICASSO* (middle) system, and *VISI* system (right). 35
2.31 Screenshots of *Search Google by Drawing* (left) and search results of a tree query (right). 36
2.32 Screenshots of the user interfaces of *Retrievr* (left) and *GazoPa* (right). 36
2.33 Screenshots of the colour picker used in *QBIC* (left) and *digiKam* (right). 37
2.34 Screenshot of the colour picker and canvas used in *imgSeek* (left) and Jacobs et al.'s (1995) [69] system (right). 37
2.35 Screenshots of Apps that use multiple colour pickers. Left-to-right: *PaintJoy*, *InfiniteDesign*, *ArtFlow*-continuous picker and *ArtFlow*-discrete picker. 39
2.36 Screenshots of the discrete colour palette used by *SketchBookX*, *Spline*, *PaintFree*, and *ArtFlow*. 39
2.37 Screenshots of three discrete colour palettes out of 16 colour palettes used in *MagicDoodle*. 40
2.38 Screenshots of *PixlrExpress* (15 shades of one hue), *WatercolorPencil Lite* (32 shades of one hue). 40
2.39 Screenshot of the colour pickers that consider colour context. Left-to-right: *FingerColors*, *AirBrush*, and *LomoPaint*. 41
2.40 A survey result of 115 responses on the frequency of drawing. . 42
2.41 The setups used for colour matching in Henry et al. [64] and Douglas & Kirkpatrick [39]. 43
2.42 The sliders with feedback (left) and without feedback (right) used in Douglas&Kirkpatrick's study [39]. 46
2.43 Five colour selector interfaces used in Kok's study [76]. 46
2.44 Screenshots with target colour (left) and colour matching process on *ColorFingers* (right) (taken from [51]). 47
2.45 The user interfaces of *Vind(x)* system (taken from [139]). 48
2.46 Screenshot of the user interface of ETM system (taken from [35]). 54

2.47 Queries of *MindFinder* system: sketch + tags (left), sketch + tag + colour (right) (taken from [17]). 54
2.48 Three sketch styles: black-white line sketch, colour line sketch, and colour blob sketch (taken from [16]). 54
2.49 The icon-based query formulation interface in Lew [84] (left), and a landscape with its icon-based query in Hayashi et al. [61] (right). 55
2.50 The pre-defined shapes of the *VISI* system (left, taken from [65]) and the sketch tool in the *CHROMA* system (right, taken from [80]). 55
2.51 Screenshots of *VisualSEEK* (left) and *WebSEEK* (right). 56
2.52 Screenshots of iPad user interface of Tseng et al.'s system [131] (left) and *SketchMatch* Windows App [125] (the rest). 57
2.53 Spatial distribution of objects on canvas. Red indicates the highest and dark blue shows the lowest density of objects (taken from [11]). 58
2.54 Screenshots of *InfinitDesign* (left) and *CanvasLite* (right). 59
2.55 Screenshots of *Sketch*. 60
2.56 12 brand logos studied by *Monochrom*. 63

3.1 The Client-Server architecture of our QbD mobile application. . 68
3.2 The set of weights used in our FMIQ implementation (taken from [69]). 69
3.3 An example of portrait colour feature files stored in our Server. . 70
3.4 Screenshot of drawing tools of our early QbD prototype. Left to right: *brush, colour picker, eraser, undo, redo, trash*, and *search*. 72
3.5 Screenshots of brush/background colour picker interface, work flow, and result-displaying user interfaces of preliminary Query-by-Colour-Drawing App. 73
3.6 Colour drawings from four human-factor experts (drawings in one column from one person) and the original paintings (last column). 75
3.7 The 'horse' drawings from experts in the focus group study. . . 77
3.8 Screenshots of the initial (left) and the refined (right) drawing tools with self-explanatory icons. From left to right: *brush*, *brush colour picker*, *eraser*, *canvas colour picker*, *undo*, *redo*, *trash*, and *search*. 79
3.9 Screenshots of the initial (left) and the refined (right) colour picker interface. 81
3.10 Screenshots of the final interfaces for a portrait query and result displaying. 81

4.1 Distribution of the painting preference among realism, abstract or others. . . . 87
4.2 Distribution of the painting preference for different themes. . . . 88
4.3 Distributions of the painting preference for different subject matter. . . . 88
4.4 The stimulus used in practice session. . . . 89
4.5 The setting of our experiments and the *Ziggi USB* camera (right). 91
4.6 One unit of the Questionnaire Q2. . . . 98
4.7 Questionnaire Q3. . . . 99
4.8 One unit of the Questionnaire Q4. . . . 99
4.9 Questionnaire Q5. . . . 100
4.10 Differentiation of background and foreground in a portrait drawing. . . . 103
4.11 The examples of coding the 'background' and 'foreground'. . . . 103

5.1 Screenshots of System A's background colour palette for portrait painting search in Experiment 1. . . . 111
5.2 Screenshots of System A's brush colour palette for portrait painting search in Experiment 1. . . . 112
5.3 Screeshots of System C's brush colour palettes (256 gradient colour shades) in Experiment 1. . . . 113
5.4 Screenshot of System C's background colour picker. . . . 114
5.5 Screenshots of System B's brush colour picker: H-SV colour square, RGB slider, and Hex value input box. . . . 114
5.6 Screeshot of System B's background colour picker. . . . 115
5.7 The stimuli with their names used in Experiment 1. . . . 115
5.8 Main effects of system design (left) and gender (right) on the time of creating a query. Error bars are the standard error of the mean. . . . 120
5.9 Main effects of system design (left) and gender (right) on the number of colours depicted in a query. Error bars are the standard error of the mean. . . . 121
5.10 Main effects of system design (left) and gender (right) on the number of objects depicted in a query. Error bars are the standard error of the mean. . . . 122
5.11 Proportions of Top21 query and Top3 query made by using each of three systems. . . . 123
5.12 Boxplot of the query's colour accuracy of the '*Lady*' queries drawn by using System A, B, and C. . . . 124
5.13 The boxplot of the foreground colour accuracy ratings of the '*Lady*' queries. . . . 125
5.14 The boxplot (left) and the cumulative percent (right) of the background colour accuracy ratings of the '*Lady*' queries. . . . 126

5.15 Distribution of participant's preference of brush colour picker. . 128
5.16 Distribution of participant's preference of background colour picker. 129
5.17 Example of drawing a contour and filling colours of the '*Lady*' queries. 132
5.18 Participants' ratings on seven attributes of our QbD App. 136

6.1 Screenshot of the query formulation tool taken from Town and Sinclair [128]. 147
6.2 Welcome screen of our QbD App used in Experiment 2. 150
6.3 Screenshots of System A's background colour palette for portrait painting search in Experiment 2. 151
6.4 Screenshots of System A's background colour palette for landscape painting search in Experiment 2. 151
6.5 Screenshots of System A's brush colour palette for portrait painting search in Experiment 2. 152
6.6 Screenshots of System A's brush colour palette for landscape painting search in Experiment 2. 152
6.7 Screenshot of prototype of System Texture (left) and the repetition mechanism of texture brush (right) (taken from *tutsplus* website). 153
6.8 Screenshots of texture palettes of Wall, Skin, Hair, Hats, Sky, Coat, Dress, Accessories in system '*Portrait*'. 154
6.9 Screenshots of texture palettes of the Sky, Sunset, Mountain, Lake, Tree, Grass, Ground, and Other objects in system '*Landscape*'. 155
6.10 The stimuli with names used in Experiment 2. (A)-(C) are used for T1 tasks, (D)-(G) are the target painting for T2-portrait tasks, (H)-(O) are used for T3 tasks. 157
6.11 Percentage of Top21-query and Top51-query made on each system for the copying-tasks in Ex1 & Ex2. 161
6.12 Percentage of Top21-query and Top51-query among queries for three types of tasks in Ex1 & Ex2. 165
6.13 Distribution of participant brush type preference with respect to painting type. 168
6.14 Distributions of queries made by participants from each of 20 countries. 169
6.15 Distributions of queries made by participants from each of 20 countries. 171
6.16 Distribution of participants' agreement ratings on Statement 1 (left) and the effects of drawing skill (right). 172
6.17 Distribution of participants' agreement ratings on Statement 2. . 173

6.18 Distribution of participants' agreement ratings on Statement 3 (left) and the effects of task type (right). 173
6.19 Distribution of participants' agreement ratings on Statement 4 (left) and the effects of task type (right). 174
6.20 Distribution of participants' agreement ratings on Statement 5 (left) and the effects of task type (right). 175
6.21 Frequency of each of four systems being chosen under repeated system selection condition. 176
6.22 Frequency of each seasonal scene query made by using four systems under single system selection condition. 177
6.23 Examples of query with side view: P3 query from Participant 38, 55, 66. 180
6.24 Examples of mixing drawn background with a background colour on queries for T3 tasks. 182
6.25 Examples of the complex background of query for task P1 (by female, Chinese Undergraduate), P2 (by female, Chinese Undergraduate), and P4 (by male, English PhD student). 182
6.26 Examples of landscape queries for T3 tasks made by using brush colours only. 183
6.27 Examples of the seasonal scene queries, top-bottom rows: spring, summer, autumn, winter. 184
6.28 Queries for four seasonal scenes from Participant 51 (male, Chinese, Engineering PhD student). 186
6.29 Summer queries from Participant 67 (female, Indian, Plant & Crop PhD student) and Participant 65 (female, English, HCI research assistant). 187
6.30 Four season scenery queries with one row from one participant. Top-Bottom: Participant 56 (male, Chinese CS PhD student), Participant 58 (female, Malaysian CS PhD student) and Participant 62 (male, Arabic IT PhD student). 188
6.31 Summer query from Participant 70 and two self-selected most-relevant images. 189
6.32 Autumn query and two self-selected most-relevant images from Participant 64. 189
6.33 The '**green tree**' in Spring query and self-selected match from Participant 70 (left two); The Autumn query with '**red leaves**' and self-selected match from Participant 58 (right two). 190
6.34 A Spring query and self-selected match (left), a Summer query and self-selected match (right) from Participant 72. 190
6.35 An Autumn query and self-selected match (left), a Winter query and self-selected match (right) from Participant 65. 190
6.36 A Summer query and self-selected match from Participant 68. . 191

6.37 Autumn query and self-selected match (left), Winter query and self-selected match (right) from Participant 60. 191
6.38 Examples of queries roughly drawn by using texture brush. . . . 196

7.1 Screenshot of of ***H-SVsquare*** colour picker interface. 206
7.2 Screenshot of of ***S-HVsquare*** colour picker interface. 207
7.3 Screenshot of of ***V-HSsquare*** colour picker interface. 207
7.4 Screenshot of of ***V-HSwheel*** colour picker interface. 208
7.5 Screenshot of of ***Vring-HSwheel*** colour picker interface. 209
7.6 The stimuli with names used in Experiment 3. (A)-(B) are used for T1 tasks, (C)-(E) are the target painting for T2-portrait tasks. 210
7.7 Main effects of task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean. 213
7.8 Main effects of task type and drawing skill on the number of colours depicted in a query. Error bars are the standard error of the mean. 214
7.9 Percentages of the Top51 query and non-Top51 query with respect to colour picker (systems) design, task type, painting type, and drawing skill. 215
7.10 The observed cumulative colour accuracy ratings of queries made by using three *bar-square* colour pickers. 216
7.11 The box-plot of participants' 7-point preference ranking scales for three *bar-square* colour pickers. 218
7.12 Proportion of colour picker selected as the '*best-to-use*' and '*worst-to-use*' *bar-square* colour picker. 219
7.13 Main effects of painting type, task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean. 220
7.14 Main effects of painting type and task type on the number of colours depicted in a query. Error bars are the standard error of the mean. 221
7.15 Main effect of task type, painting type and drawing skill on the number of objects depicted in a query. Error bars are the standard error of the mean. 222
7.16 Drawings created by participant with higher drawing skills (DS, 1=lowest, 5=highest) in Experiment 3. 222
7.17 Percentages of the Top51 query and non-Top51 query in terms of two wheel-based colour pickers, task type, painting type, and drawing skill. 223
7.18 The observed cumulative colour accuracy ratings of queries drawn via two wheel-based colour pickers. 225
7.19 The symmetry distribution of preference difference between *V-HSwheel* and *Vring-HSwheel*. 226

7.20 Proportion of colour picker being selected as the best-to-use and worst-to-use 2D-wheel colour picker. 227
7.21 Main effects of system, task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean. 228
7.22 Main effects of painting type and task type on the number of colours depicted in a query. Error bars are the standard error of the mean. 228
7.23 Percentages of the Top51 query and non-Top51 query in terms of two bar-based colour pickers, task type, painting type, and drawing skill. 230
7.24 The observed cumulative colour accuracy ratings of queries made by using *V-HSsquare* and *V-HSwheel*. 231
7.25 Almost symmetry distribution of preference difference between *V-HSwheel* and *V-HSsquare*. 232

8.1 Screenshots of the brush colour picker (left) and background colour picker (right) of QbD system used for the longitudinal memory drawing study. 240
8.2 The stimuli with names used in Experiment 4: (A)-(D) are for 15-min delay session; (E)-(F) for 1-week delay; (G)-(H) for 1-month delay). 241
8.3 Distribution of participants' self-rated drawing skill (1=worst, 5=best). 244
8.4 Main effects of the study trial (left) and drawing skill (right) on the time of drawing a query. Error bars are the 95% confidence interval of the mean. 245
8.5 Main effects of the study trial (left) and painting type (right) on the number of colours depicted in a query. Error bars are the 95% confidence interval of the mean. 246
8.6 Main effects of the study trial (top), painting type (bottom-left) and experience with QbD (bottom-right) on the number of objects depicted in a query. Error bars are the 95% confidence intervals of the mean. 248
8.7 Proportions of Top51 queries in all queries drawn during three study trials (top), and in terms of gender (bottom-left) and the experience with our QbD App (bottom-right). 249
8.8 Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest). 251
8.9 Plots of cumulative percentage of participants' self-rated similarity ratings regarding study trials (left) and participant's experience with our QbD App (right). 251
8.10 Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest). 252

8.11 Plot of cumulative percentage of colour accuracy ratings from three study trials. 253
8.12 Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest). 254
8.13 Plots of the cumulative percentage of participants' self-rated similarity ratings regarding study trials (left) and participant's experience with our QbD App (right). 254
8.14 Participant 4's '*Flowers*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial). . 255
8.15 Participant 40's '*Flowers*' queries during first two study sessions (left-to-right: 15-min delay, and 1-week delay trial). 256
8.16 Participant 43's '*Pink*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial). . 256
8.17 Participant 31's '*Flowers*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial). . 257
8.18 The '*Seashore*' queries of Participant 28 (experienced, male) (top row) and Participant 81 (novice, female) (bottom row) during 1-week delay (left) and 1-month delay session (right). 258
8.19 The '*Flowers*' queries of Participant 81 during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial). 259
8.20 The '*Seashore*' queries from Participant 75 during 1-week delay trial (left) and Participant 80 during 1-month delay trial (right). . 259
8.21 The '*Flowers*' queries of Participant 74 (top row) and Participant 79 (bottom row) during 15-min delay (left), 1-week delay (middle) and 1-month delay (right) session. 260
8.22 Queries of the '*Seashore*' from Participant 76 (novice, male) during 1-week delay (left) and 1-month delay (right). 260
8.23 Queries of the '*Countryside*' from Participant 28 (left), Participant 76 (middle), and Participant 77 (right). 261

B.1 Colour Value for k-means Background Colour Palette used in Experiment 1 291
B.2 Colour Value for k-means Background Colour Palette for portrait painting search in Experiment 2 292
B.3 Colour Value for k-means Background Colour Palette for landscape painting search in Experiment 2 293

List of Tables

2.1 Distributions of colour pickers used in reviewed Android Apps on *GooglePlay*. 38

5.1 Distribution of 256 discrete colours in 12 colour categories. . . . 112
5.2 Results of GEE models exploring effects of system design on the participants' choices of each usability feature. 127
5.3 Results of GEE models exploring effects of colour picker type on the participants' choices of each usability feature (n.s.: not significant). 128
5.4 Results of GEE models analysing the effects of colour picker type on participants' preference choices (n.s.: not significant). . 130
5.5 Distribution of the first interaction actions for different stimuli. . 130
5.6 Frequency of 11 types of refinement action for the '*Lady*' task. . 134
5.7 Summary of colour picker type effects on measures for the '*Lady*' queries (n.s.: not significant). 138
5.8 Summary of System effects on measures for the '*Lady*' queries (n.s.: not significant). 141

6.1 Distribution of 256 *k-mean* brush colours for portrait painting and landscape painting search in seven colour categories. 151
6.2 Results of GEE models for analysing system effects on all copying tasks from Ex1 & Ex2 (n.s.: not significant). 161
6.3 Results of GEE models for colour picker type effects on all copying tasks from Ex1 & Ex2 (n.s.: not significant). 163
6.4 Results of GEE models for analysing system effects on all queries from Ex1 & Ex2 (n.s.: not significant). 163
6.5 Results of GEE models for analysing colour picker type effects on all tasks from Ex1 & Ex2 (n.s.: not significant). 164
6.6 Summary of task type effects on measures for all queries from Ex1 & Ex2. 164
6.7 Summary of painting type effects on measures for all queries from Ex1 & Ex2 (n.s.: not significant). 167
6.8 Summary of brush type effects on measures for all queries from Ex1 & Ex2 (n.s.: not significant). 167

6.9 Significant ethnic group effects on measures for T2 queries, T3 queries, and T2 & T3 queries from Ex1 & Ex2. 170
6.10 Comparisons of system effects on the '*Lady*' queries from Ex1 (Effect-Lady), all copying-task queries (Effect-T1) and all tasks on colour systems (Effect-T1∼T3) from Ex1 & Ex2 (n.s.: not significant, '-': not available). 192
6.11 Comparisons of colour picker type effects on copying tasks from Ex1 and on all copying tasks from Ex1 & Ex2 (n.s.: not significant). 193

7.1 Results of Friedman tests, Median and IQR of 7-point ratings of usability of three HSV-combination colour pickers. 217
7.2 Results of Wilcoxon tests, Median and IQR of 7-point ratings of usability of *Vring-HSwheel* and *V-HSwheel* colour picker. 225
7.3 Results of Wilcoxon tests, Median and IQR of 7-point ratings of usability of *V-HSsquare* and *V-HSwheel* colour picker. 231
7.4 Summary of the colour parameter combination effects (n.s.: not significant). 233
7.5 Summary of effects of the shape of 1D input area (n.s.: not significant). 234
7.6 Summary of effects of the shape of 2D input area (n.s. =not significant). 235

8.1 Summary of the delayed trial effects (n.s. = not significant). S1=15min-delay session, S2=1week-delay session, S3=1month-delay session. 262
8.2 Summary of significant effects of the experience with QbD App. 263

To my family...

Chapter 1

Introduction

With the advent and proliferation of personal touchscreen mobile devices, the study of drawing and its role in image search have entered a new frontier. The overall aim of this thesis is to explore the key factors affecting the usability and task performance of the Query-by-Drawing (QbD) painting search application on mobile devices. A better understanding of these effects and user interactions could improve the user interface design for QbD-related application.

In this chapter, the research background is firstly introduced, followed by highlighting our motivations and objectives (Section 1.2) and the research questions (Section 1.3) generated from the background literature. The major contributions are briefly presented in Section 1.4. Finally, Section 1.5 will outline the structure of the whole thesis.

1.1 Background

Current commercial image search engines are mainly keyword-based, several limitations exist. First, the text-based image search technology requires a rather

laborious and tedious effort to annotate images (especially for large image collections). Second, users might use different words to describe an image, even the same person would tag different keywords for the same image at different times. Moreover, sometimes it might not be sufficient or ideal to describe a user's search ideas in several keywords.

To find an image by using a current search engine, users have to specify a combination of words as a query. Cunningham et al. [28] analysed and categorised 404 authentic visual art queries posted to *Google Answers 'ask an expert'* system, and found that most queries contained bibliographic metadata. For example, the most frequently used metadata contains the artist name, the date of creation, and the artwork title, followed by the content description of the art piece, e.g., using colour words to describe objects on the artwork ('*wearing purple hat*') and the overall impression ('*blue-greenish canvas*'). However, due to the ambiguity of natural language, most painting images or artwork collections on websites may not contain the tags precisely matching user's keywords. In case users do not know the title and artist's name of an art piece, the textual content description may be highly variable because of different individual's semantic interpretation and visual emphasis [21]. As Pu [109] found, failed queries of image search usually contain conceptual features such as emotive response. Moreover, it is hard for people to describe a painting in several keywords when their search intent is vague, and some paintings (e.g., abstract paintings) may be difficult to define verbally [130]. Hence, *Content-based Image Retrieval* (CBIR) technology is, in particular, a useful supplement to the keyword-based query.

Content-based Image Retrieval has been proposed since the early 1990s [113]. There has been a lot of research into CBIR techniques (see [30, 105] for review), which adopt features of colour, shape, texture or combinations as the 'signature' of an image and search results are based on the feature comparison between a query and database. However, most CBIR systems have been developed for academic research purposes rather than fully developed end-user

systems in widespread practical use. The reasons for this are twofold. First, an image query is required by CBIR systems to start a search and the search results highly depend on this initial query image [140], yet currently only a few CBIR studies focus on user interfaces of CBIR systems to facilitate end users to express the visual query. Second, the utility of a CBIR system is limited to the type of query and database images in the specific area [137]. It is suggested that making the initial query representative enough of user's search intent is vital to avoid the *Page Zero problem* [74].

In this thesis, we limit ourselves to painting search. For instance, people might aim to search for a fairly uncommon painting and they only have a rough idea of what the painting looks like. In this case, drawing is a more concise way than verbal language to express search intent. McDonald et al. [92] suggested that sketch-based image search could be used as a complement to text-based image search. According to Banfi [6], drawing is the most flexible way for users to present their search ideas, and almost everybody is capable of performing the simple activity of drawing [33]. A picture could paint a thousand words [127].

With the rapid growth of the use of touchscreen mobile phones or tablets recently, there are lots of potential opportunities to improve CBIR system by facilitating users to draw a query. Because touchscreen devices offer a richer set of natural user interactions, e.g., drawing by using fingers or stylus, which are more intuitive and direct for inputting a visual query than using a mouse on the computer [136]. McDonald and Tait [93] revealed that users tend to put more focus on colour rather than the overall layout of the scene depicted in an image. Moreover, on small screen phones, it might be simpler and more feasible to draw rough colour strokes for the most prominent colour patches rather than the complete details of objects [69]. Therefore, a colour selector is necessary for a mobile Query-by-Drawing (QbD) system.

Zhang [152] proposed that a user-friendly and artistic user interface would attract and retain users. To the best of our knowledge, although an increasing number of innovative QbD techniques used for mobile devices have been proposed in the academic literature [75, 125, 130], little work has been focused purely on improving the user experience of image search through exploring and designing user-friendly user interfaces. Only a few guidelines and heuristics on the methodology used for exploring the QbD user interface on a computer, which could be selectively employed in the QbD mobile interface study, are found in the literature [66, 92, 93]. Moreover, the colour selector interface design is still an active area of research [42, 98], and there are thousands of colour selector designs on the current mobile application market. Prior work mainly examined simple colour selectors (e.g., colour sliders) based on a colour matching experiment [39, 64, 133]. However, the colour matching task is out of the context of real-world drawing scenario, in which choosing one colour might be influenced by the surrounding or background colours. A review of the literature also reveals that there are a few common principles of colour selector interface design for drawing a query on mobile devices.

1.2 Motivation and Objectives

Inspired by the literature, rather than investigating the image retrieval algorithm itself, we believe that a better understanding of factors influencing the usability of mobile Query-by-Drawing (QbD) systems and the human drawing process could help researchers to design more user-friendly and intuitive drawing interfaces, on which users could produce a qualified query to specify their search intent. Therefore, we narrow down our research to painting search on mobile phone based on an existing QbD algorithm. Previous work on understanding image search strategies has largely been focused on classifying the kinds of queries produced by professionals.

Designing intuitive Query-by-Drawing (QbD) user interfaces on mobile phone is one of our primary goals. The potential factors examined in this thesis consist of the colour picker type and their combinations, the visual design attributes of continuous colour picker, the brush type, the task type, and user characteristics. We are also interested in whether the approach of drawing rough colour sketches as a query is viable for real use, e.g., search for painting in memory. Specifically, the main objectives of this PhD work are:

- To evaluate the query-drawing process by examining how people use different drawing interfaces to draw queries for different types of tasks, and to collect users' opinions about the user interfaces and the results returned by the system, as well as suggestions for improving the design of our QbD application.
- To explore the effects of combinations of continuous and discrete colour pickers on task performance, system usability and user preference, and to compare the difference between continuous colour picker and discrete colour pickers.
- To determine the effects of task type, painting type, brush type, and user characteristics such as gender, ethnicity and drawing skill on task performance, usability and user preference, if any.
- To see and understand how different task types would lead participants to use one system over another.
- To explore the effects of attributes of continuous colour pickers: the combination of three colour parameters and the shape of input area on task performance, usability and user preference of colour picker interface.
- To determine whether or not the user could draw rough colour sketches from memory of previously-seen paintings over different delayed periods, and to determine the factors affecting the memory drawing over time.

1.3 Research Questions

The objectives lead to the following research questions (RQ):

- **RQ1** - How do continuous colour picker and discrete colour picker and their combinations affect users' task performance, preference, and the usability of a QbD painting search application on mobile devices? (addressed in Chapter 5 and Chapter 6)
- **RQ2** - How do users use different user interfaces to draw queries for different types of tasks? (addressed in Chapter 5, Chapter 6, Chapter 8) What do users feel about the painting search by drawing colour sketches on mobile phones? (addresses in Chapter 5)
- **RQ3** - How do task type, painting type and brush type affect the task performance, usability and user preference of a QbD mobile application? (addressed in Chapter 6)
- **RQ4** - How do the combinations and visual representations of three colour parameters of 1D-2D continuous colour pickers influence users' task performance, preference and usability of a QbD mobile application? (addressed in Chapter 7)
- **RQ5** - Do people have the ability to draw a query for the painting viewed (15 minutes ago, 1 week ago, and 1 month ago) previously? What factors and how they affect people's drawings of their painting memory over time? (addressed in Chapter 8)

1.4 Contributions

The thesis begins with a comprehensive literature review and a survey of 189 existing Android applications (Apps) adopting colour picker on *Google Play*,

to get an overview of the current state of Query-by-Drawing and user interface design of mobile drawing tools (Section 2.2.3.2).

In order to answer these research questions, we have developed a Client-Server based Query-by-Drawing (QbD) mobile application to allow users to search for paintings by drawing colour sketches (Chapter 3). We build several painting databases which are publicly accessible from the *Tate collection* and the *BBC Your Paintings* website (Section 3.2.3). A focus group is carried out to evaluate the usability of a preliminary prototype of our QbD App (Section 3.3.1), and its findings are used to redesign both general drawing interfaces and the colour picker interfaces.

Four experiments involving 123 participants in total are conducted with different purposes and different colour picker interfaces. These experiments result in a better understanding of the patterns of the query drawing process and the effects of colour picker design, brush design, task type, painting type and user characteristics at group levels, which would help researchers to capture a comprehensive picture of the use of colour picker and the guidelines for designing user-friendly QbD mobile applications. The methods for our experimental design, data collection and data analysis also contribute to the literature. For example, to make the experimental tasks more representative and realistic, we have conducted an online survey about people's painting preference to decide the stimulus type used in our experiments (Section 4.3.1), from which we decided to select the stimuli from realistic landscape paintings and portrait paintings.

An outline of the core contributions of this thesis is as follows:

- A systematic review of colour selectors used in current published CBIR systems;
- A survey about the current mobile phone use and the colour pickers used in various recent Android mobile applications;

- A new way of categorising colour pickers is proposed: '1D-1D-1D', '1D-2D', and '3D' colour picker;
- An implementation of a Client-Server based Query-by-Drawing (QbD) mobile application for painting search, which could be used as a platform for testing other QbD algorithms and future exploratory research;
- A survey on painting preferences and the rationales to make choices of stimulus are formulated;
- A real-world painting database is built;
- An experimental comparison of the effects of discrete colour picker and continuous colour picker;
- An empirical comparison of the use of plain colour brush and textural brush;
- An experiment exploring the effects of the visual design attributes of 1D-2D (1 dimensional-2 dimensional) continuous colour picker, including the colour parameter combinations, the shape of 1D input area, and the shape of 2D input area on the usability of QbD application and task performance;
- An experiment examining users' drawings from memory with three delayed periods 15-minute delay, 1-week delay, and 1-month delay;
- An exploratory observation of users' search behaviour and drawing patterns on four types of tasks drawing with a target in sight, drawing from the description, drawing immediately after viewing a target, and drawing from memory after long-term delayed viewing a target;
- Examining users' query refinement of using our QbD App for copying-task;

- A comprehensive understanding of the users' result relevance judgement strategy for target unspecific tasks and the memorisation strategy used for target-absent tasks;
- A combined use of subjective and objective measures, descriptive statistics and inferential statistics, and qualitative and quantitative methods;
- Novel methodologies for experiment design, data collection, and data analysis contribute to the literature, e.g., a new protocol for drawing accuracy ratings.

To sum up, the novelty of this work is the development of a mobile QbD application (Chapter 3) with a real-world painting database, which allows users to draw rough colour sketches and to search for different kinds of tasks. In addition, a series of colour picker interfaces and texture palettes have been implemented and examined, including the comparison between discrete and continuous colour pickers, the comparison within continuous colour pickers with different visual design attributes, and the comparison of plain colour and texture brushes. To the best of our knowledge, this work is the first experimental investigation in exploring the effects of the colour picker used in real-world CBIR image search tasks by drawing on mobile phones.

1.5 Organization of the Thesis

Followed by the introduction of this PhD research in Chapter 1:

- In Chapter 2, we review several related topics including colour models, colour pickers, colour selection in CBIR system, Query-by-Drawing,

colour memory, drawing from memory, and search and drawing strategy. We briefly discuss a series of general colour pickers and the existing colour pickers used in CBIR systems and current mobile applications. To get an overall concept of current drawing study, both psychological study and computer vision study in drawing, especially Query-by-Drawing paradigm, are outlined. In addition, methodologies used in the research of colour memory, memory drawing, search strategies and drawing patterns are reviewed to get an overall picture of related research to facilitate our experimental design.

- In Chapter 3, we introduce a full-scale prototype Query-by-Drawing (QbD) mobile application with real oil paintings as image database. Our QbD App is based on the Client-Server architecture, which separates the user interaction from back-end image retrieval algorithm. The QbD App is emphasised on the user interface design of different colour pickers. A focus group is presented, aiming for evaluating the usability of general user interfaces and feasibility of Query-by-Line-Drawing technique, followed by describing improved user interfaces of drawing tools and colour picker in the final section.

- In Chapter 4, we present the methodology used for experimental design, data collection and data analysis. Methods such as questionnaire, interview and think-aloud protocol are briefly introduced. Most importantly, we describe the experimental setup, the choices of stimuli, the task design, the questionnaire design, and evaluators' subjective ratings, which are generally used for all experiments. At the end of the chapter, we discuss both justifications of the methodology and the limitations of our experimental design.

- In Chapter 5, we introduce Experiment 1 which involves 37 participants to explore the effects of the continuous colour picker and discrete colour

picker and their combinations on task performance and usability of our Query-by-Drawing (QbD) application. The results of statistical analysis based on queries for copying tasks are presented. Participants' responses on the usability and the overall QbD system are discussed. We also present the participants' drawing patterns and query refinement behaviour, as well as the search strategies adopted during the search process by observing and mainly by reviewing the screen-videos.

- In Chapter 6, we introduce Experiment 2 with 38 participants, which serves as an implementation of Experiment 1 by adding tasks of searching for landscape paintings and more stimuli for three task types. Another purpose of Experiment 2 is to evaluate and compare two types of brushes - plain colour brush vs. texture brush. The refined user interfaces of general drawing tools and two new textural systems are also described. We present and discuss the statistical analyses based on the queries for copying-tasks from Ex1 and Ex2 to verify the findings from Experiment 1. The statistical analyses based on all queries produced on three colour systems from Experiment 1 and Experiment 2 aim to explore the effects of task type, painting type, and user characteristics; and the statistical analyses of queries made by using plain colour brushes and texture brushes examine the effects of brush type. Most importantly, we investigate the participants' attitudes, the effects of task type on system choice, the effects of ethnic group and the drawing pattern when performing real-world tasks.

- In Chapter 7, we describe Experiment 3 which examines how visual design attributes of a 1D-2D continuous colour picker affect the usability of a QbD image search App. Specifically, the combinations of three colour parameters and the representation of 1D input area and 2D input area are explored and built into five continuous colour pickers, which are evaluated by 41 participants. Statistical analyses are made mainly for exploring the

effects of these three attributes on task performance and user perceived usability and preference of different colour picker interfaces.

- In Chapter 8, a longitudinal memory drawing study (Experiment 4) is presented, with the purpose of exploring the viability of the Query-by-Drawing paradigm for a real-life search scenario. Drawn queries at three time points after viewing stimuli are explored: 15-minute delay, 1-week delay, 1-month delay. 17 participants and 8 painting stimuli are adopted. The data analyses mainly focus on how these delayed periods, painting type and user characteristics such as gender, drawing skill, and the experience with our QbD App, affect the task performance, the quality of drawing and users' confidence. Memorising strategies and drawing behaviour are also discussed at the end of this chapter.

- Finally, Chapter 9 discusses the overall achievements of this PhD project by reviewing our contributions and suggesting some future research directions.

Chapter 2

Review of Related Work and Methods

2.1 Introduction

Before investigating factors affecting the usability of the Query-by-Drawing system, this chapter reviews some related work in the fields covering the key topics: colour models, colour pickers, colour selection study, Query-by-Drawing paradigm, drawing-related studies, and drawing from memory.

In Section 2.2, we introduce a series of topics related to colour selection. First, we briefly review colour models (Section 2.2.1) and common colour picker shapes (Section 2.2.2.1) and categories (Section 2.2.2.3). Then, we present the status of colour pickers used in existing content-based image retrieval systems (Section 2.2.3.1) and survey results of colour pickers built on current mobile applications (Section 2.2.3.2). Additionally, previous research on essential elements of a colour picker: the colour model and visual design, as well as the typical experimental design used in exploring the colour selection strategy and the colour selection in existing CBIR system are presented in Section 2.3.

We introduce an overview of previous studies on drawing in Section 2.4, including the factors affecting drawing quality from a psychological perspective (Section 2.4.1) and the current development of the *Query-by-Drawing* paradigm on both computers and mobile devices (Section 2.4.2). To gain the rationales of our experimental design and result analysis, popular experiments with regard to colour memory and memory drawing are presented in Section 2.5. The literature on related works of selection strategy (Section 2.3.4) and drawing behaviour (Section 2.4.3) are also presented.

2.2 Colour Selection and Colour Picker

Colour is commonly used in designing web pages and user interfaces, document editing and data visualisation [91]. Colour is also one of the dominant features in Content-based Image Retrieval. It is a straightforward way for users to use colour to describe the visual content of an image [141]. As McCamy [91] proposed, under normal conditions 1,875,000 colours could be discernable and 7.5 million colours under good conditions. Hence, a colour selecting tool is necessary for the user to access colours on digital devices.

From the renaissance until now, people attempt to define colours and to arrange colour in the logic order based on scientific observations from experiments. Some researchers (or artists) focussed on the psychological effect of colour. With the advent of digital technology, to represent colours in a more intuitive way, many explorations have been made to create a 'natural' colour model (or colour space) and an 'ideal' colour selector. As the emergence of a wide variety of software involving colour selection, such as a diagram-creating applications, more and more non-professional users start to use colour pickers.

2.2.1 Colour Model

A colour model is a mathematical representation of colour with normally three or four numeric values; a colour space is a specific implementation of a colour model and has a specific range of colours. Each colour is represented by a single point [57] in a three-dimensional (3D) colour space. Different colour spaces based on a colour model are available for different applications [48]. For example, the RGB (*Red-Green-Blue*) colour model is used by a number of colour spaces [104]: sRGB, Adobe RGB, Apple RGB and CIE RGB etc. The term 'colour model' and 'colour space' are often used interchangeably. In this thesis, these two terms are the same unless otherwise noted.

There are a number of colour models devised for specific tasks or applications. Some models are proposed to help humans to select colours (e.g., Munsell system [97]), and some are designed to facilitate data processing. Many studies have explored whether some colour models are more appropriate than others [38]. However, it is still a controversial topic about which colour model is the better than others.

By comparing the procedure of selecting a target colour by using the Red-Green-Blue (RGB), Hue-Saturation-Lightness (HSL), and Natural Colour System (NCS) colour model, Murch [98] proposed that the HSL colour model requires the fewest steps for the inexperienced users. Schwarz et al. [115] reported that the RGB colour model produced quick but inaccurate matches in a comparison to the Hue-Saturation-Value (HSV) model. This is inconsistent with the results from Wells's study [144], which investigated three colour models (RGB, HSV, and Opponent channel), and concluded that there were no differences in the time and accuracy of colour selection among these three colour models. Similarly, Douglas and Kirkpatrick [39] reviewed 12 colour models and found that the colour model has a non-significant impact on the performance of colour-matching tasks.

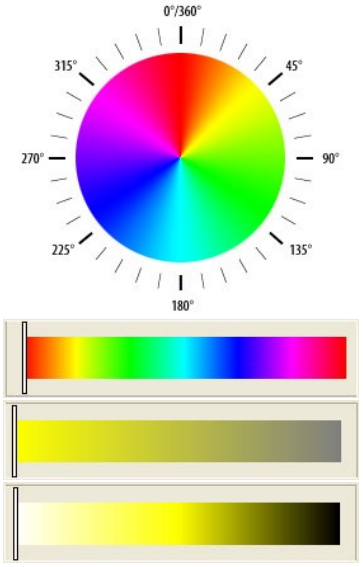


FIGURE 2.1: Representation of a *hue* wheel, a *hue* bar, a *saturation* bar, and a *value* bar.

Colour model is only one aspect of colour selection. The visual design of a colour picker that allows the user to directly manipulate the colour model must also be considered. According to Lyons et al. [89], most non-professional users who are unfamiliar with colour model might have difficulties in correlating a specific colour with a location in the 3D colour space. In this thesis we adopt the HSV (*Hue-Saturation-Value*) [120] colour model, because its concept is similar to a natural way that human think about a colour. *Hue* describes a dominant colour, while *saturation* measures the proportion of greyness in a colour [138]. The *brightness*, *intensity*, *lightness* or *luminance* are equivalent terms to the *value*. *Hue*'s value is defined with beginning and ending with red and running through all intermediary colours like *orange-yellow*, *greenish-blue*, *purple*, etc., as shown in Figure 2.1. *Saturation* runs in percent ranging from 0 (no colour *saturation*, grey) to 100 (fullest *saturation*, deepest colour). The less saturated, the more greyish a colour appears. *Value* determines the brightness from black (value 0) to fully bright (value 100). Any colour with zero brightness value is black, regardless of its hue or saturation.

2.2.2 Colour Picker: User Interface for Colour Selection

A colour picker or colour selector is a part of graphic user interfaces that allows users to select colours for use in software such as statistical software, document editors, drawing applications and web page design. The colour selector interfaces vary considerably to satisfy specific requirements. However, several common design attributes consistently appear, including a visual design of a colour palette, an input method and an underlying colour model. Similar to pigment-based colour mixing, the visual design of a colour palette defines and displays colours, and the input method is a way to allow users to mix for an expected colour, and the colour model is analogous to different kinds of pigments. To the best of our knowledge, no textbooks or articles about designing colour picker have been published.

Before considering the impact of colour picker on usability and task performance, it is necessary to explore what the existing colour pickers are and which colour pickers are prevalent among the state-of-the-art software and mobile applications.

2.2.2.1 Representation of Colour Picker

Six general categories are recognised in terms of the shapes of 1D or 2D colour palette representation: *colour swatches* (or *colour box*), *colour sliders*, *colour ring*, *colour wheel* (or *colour circle*), *colour square* (or *colour panel*), and combinations of these basic shapes.

Colour Swatch Colour swatch (also called *colour template* or *colour box*), is a representation of a discrete colour palette [146], often contains a finite number of pre-defined standard colours or user-customised colours. Each colour is

displayed separately in the form of a box or a circle, or even a hexagon (e.g., *Microsoft Office*^{TM}'s standard colour palette shown in Figure 2.2).

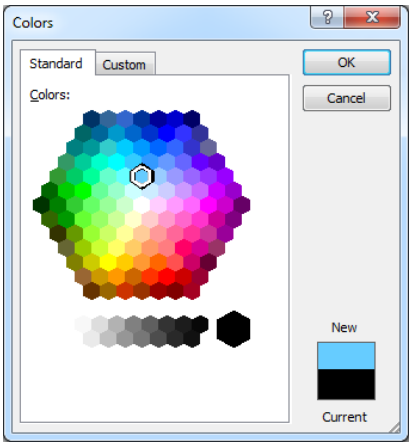


FIGURE 2.2: Screenshot of the standard colour palette in *MS Office*^{TM} Suite.

The colour sets and the number of colours are under the developers' control, which means that colour swatches could be significantly varied according to the individual's colour expertise and personal preference. The display order of colour patches in different applications is also different. For example, as illustrated in Figure 2.3, the colour swatches in *Photoshop*^{TM}, the colour palettes used in *Retrievr*[1]and in *Search Google by Drawing*[2] online application totally differ from each other.

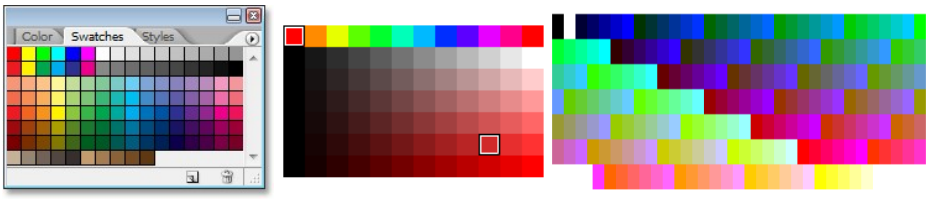


FIGURE 2.3: Colour swatches used in *Photoshop*^{TM}, *Retrievr* and *Search Google by Drawing*.

Slider Bar Slider bar is the simplest way to map one dimension in a colour model to a one-dimensional (1D) interactive colour selection tool. This can

[1] http://labs.systemone.at/retrievr/
[2] http://search-by-drawing.fullstackoptimization.com/

be applied to a set of colour parameters of any colour model, for example, the additive primary colours *Red, Green* and *Blue* could be selected by three slider bars, and the subtractive primary colours (*cyan, magenta, yellow* and *key (black)* are represented by four slider bars. Each of the RGB sliders is shaded from black (value 0) to the full intensity of a colour (value 255), and each of CMYK sliders is shaded from 0% (lighter colour) to 100% (darker colour), as the RGB and CMYK sliders of *Illustrator*^{TM} shown in Figure 2.4. When users move any slider bar, the resulting colour is mixed its value with the values of other sliders.

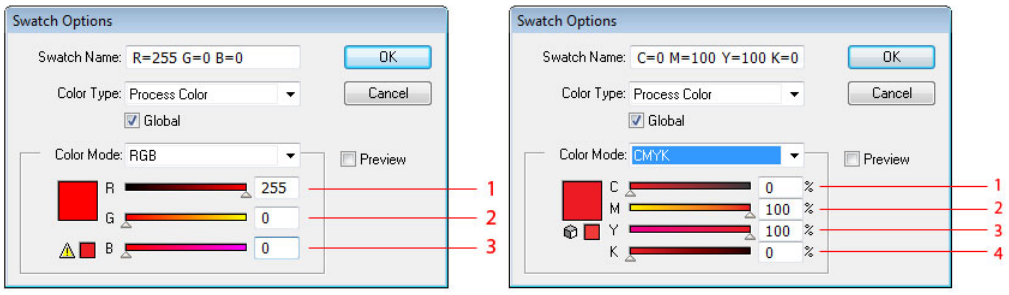


FIGURE 2.4: RGB sliders and CMYK sliders used in *Illustrator*^{TM} (taken from http://design.tutsplus.com/tutorials/an-introduction-to-illustrators-color-tools--vector-3549).

The slider bar also could manipulate the *lightness* or *saturation*, even the *hue* of a colour. Users could first drag the *hue* slider to pick a colour band, and then move the *saturation* slider to set the purity of a colour (the lower *saturation* value, the more grey a colour is) and the *value* slider to settle the brightness from black (value 0) to a fully bright colour (value 100). The HSL sliders are similar to HSV sliders except that one has a *lightness* slider and HSV has a *value* slider. Higher *lightness* values add white to a colour, whereas lower *value* values subtract white from a colour, as shown in Figure 2.5.

Colour Ring A colour ring is a circular representation of colour arrangement based on a chromatic relationship. Sir Isaac Newton first split the sunlight into seven beams: *red, orange, yellow, green, blue, indigo* and *violet*, and then he joined two ends of the colour spectrum into a colour ring [72]. After that,

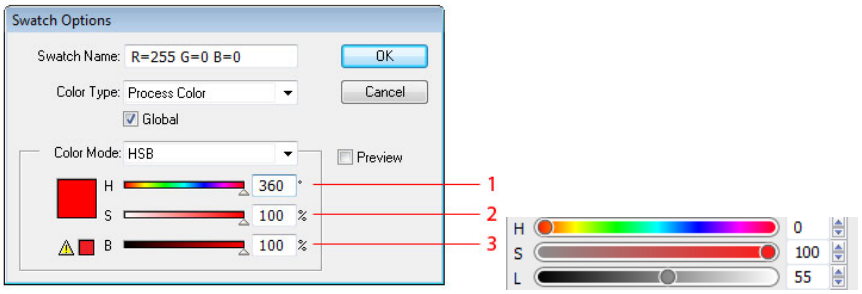


FIGURE 2.5: HSV (or HSB) sliders in *Illustrator* (left) and HSL sliders in *DrawPlus X5*.

artists and scientists studied and devised many variations of *Newton's colour ring*, serving for different purposes of mixing and selecting colours. Different colour rings can be created based on different types of paints or colour models, see Figure 2.6. All colour rings follow the same techniques and principles to represent colours: comprising of different primary and secondary colours but the exact same complimentary colours.



FIGURE 2.6: Colour rings with different colour models: RYB, RGB, and CMYK.

A colour ring is commonly used in art and design. For example, many artists and designers use the colour ring to pick harmonious colour combinations, so-called *Colour Scheme*. The current colour ring used in software and application is a continuous colour ring, which is a circular diagram of the *hue* slider. Typically, the colours in a colour ring are fully saturated colours, which are usually further refined by changing the values of their *lightness* and *saturation* in real use context, such as the colour rings shown in Figure 2.7.

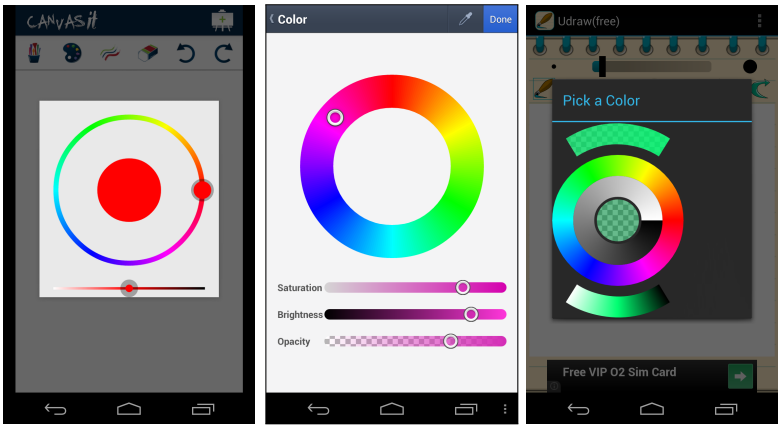


FIGURE 2.7: Screenshots of three colour rings extracted from the reviewed current mobile Apps.

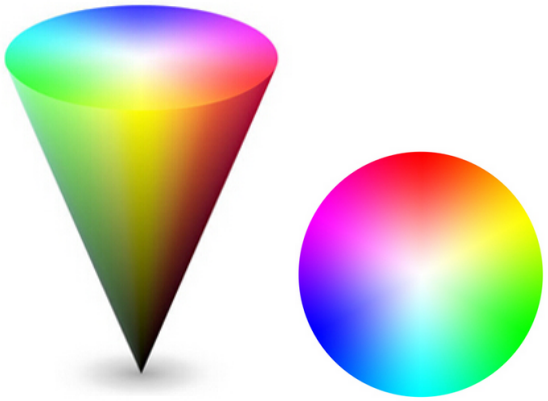


FIGURE 2.8: Representations of a HSV cone and a *hue-saturation* colour wheel.

Colour Wheel Colour wheel (also refer to as *colour circle*) is the extension of colour ring. Firstly, the colour ring is fully filled with colours; secondly, there are more colours in a colour wheel than those in a colour ring. Some researchers use the term of colour ring and colour wheel interchangeably. The colour wheel is a one-dimensional horizontal cut of the three-dimensional (3D) colour space, e.g., the *hue-saturation* based colour wheel shown in Figure 2.8, which means that only a limited number of colours in colour space are incorporated into a colour wheel. Therefore, the colour picker used in application often provides the third dimension, such as a value slider, to access more colours. We will discuss the shape-combined colour pickers in Section 2.2.2.2.

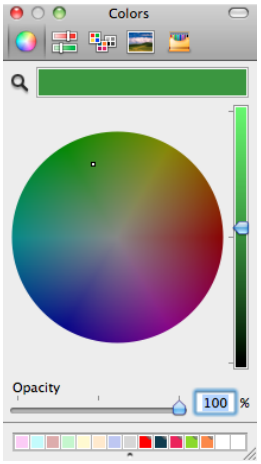


FIGURE 2.9: Screenshot of the colour picker used in *Apple's Mac OS X* system.

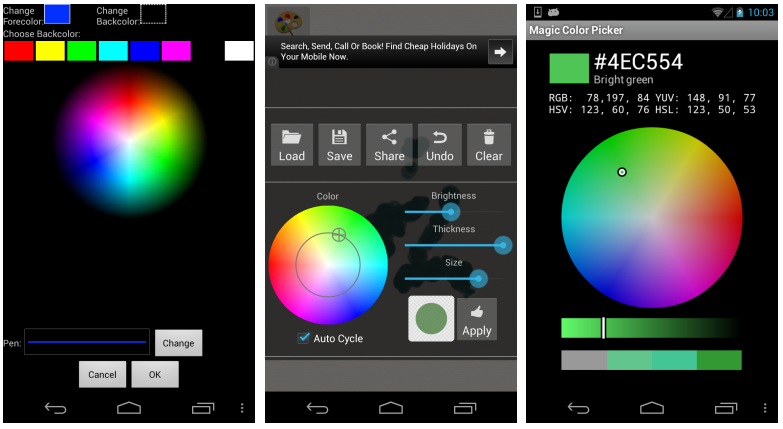


FIGURE 2.10: Screenshots of the examples of an original colour wheel (left), a colour wheel with 90 degree anticlockwise rotation (middle), and a colour wheel with 90 degree clockwise rotation (left) used in three mobile Apps.

There are many variations of the original colour wheel. One well-known colour wheel example is *Apple*'s colour picker, see Figure 2.9. Almost all *Apple* software utilises this colour wheel, e.g., *Apple's iWork* suite of *Keynote, Numbers* and *Pages*, and the built-in applications such as *TextEdit*. Some applications simply rotate the colours 90 degrees clockwise or anti-clockwise, see Figure 2.10.

Colour Square Similar to colour circle, colour square is a cutting slice of a 3D colour space except for displaying colour in a square. For example, three

RGB colour squares shown in Figure 2.11. For a 3D colour model, two dimensions of a colour are mapped to a 2D plane, and the third dimension is included as a linear slider to access more colours, see details in Section 2.2.2.2.

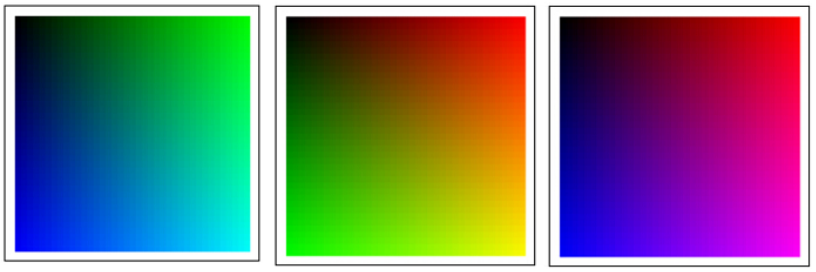


FIGURE 2.11: Representations of *green-blue, red-green*, and *red-blue* colour squares.

Figure 2.12 presents a colour square of HSL (left) and HSV (right) with *saturation*=255. *Hue* is on the horizontal axis, and *lightness* or *value* is on the vertical axis (with the maximum value of *lightness* or *value* on the top).

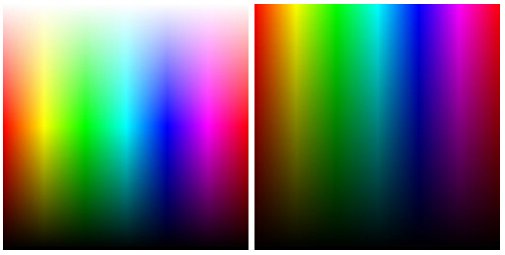


FIGURE 2.12: Representations of *hue-lightness* (left) and *hue-value* (right) colour squares, with *saturation*=255.

2.2.2.2 Colour Picker of Combined Shapes

As mentioned in the previous section, to make it possible to choose all potential colours, the actual use of the colour picker in applications combines different representations of a colour picker. The common way is to mix two dimensions into a plate which would be changed when the third dimension changes. The following parts of this section will introduce the combined shapes of colour pickers adopted in current popular software and applications.

Bar-Wheel The *Bar-Wheel* combination usually represents a *hue*-based colour model, in which a linear or circular slider bar decides the *value* or *lightness*, or *saturation*. For example, Figure 2.13 illustrates a traditional approach for the colour selection: *S-HV*, *V-HS*, *S-HL*, and *L-HS* colour wheel:

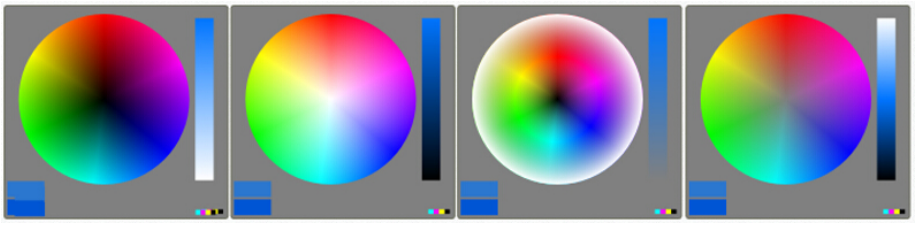


FIGURE 2.13: Representations of *S-HV*, *V-HS*, *S-HL*, and *L-HS* colour wheel. (taken from https://community.kde.org/Krita/Community_Mockups_and_Wishlist)

Taking the *V-HS* colour wheel as an example, the way to pick a colour could be: firstly selecting a *hue-saturation* value pair in the HS colour wheel, and finally adjusting the lightness in the gradient *value* bar below the HS colour wheel as seen in Figure 2.14.

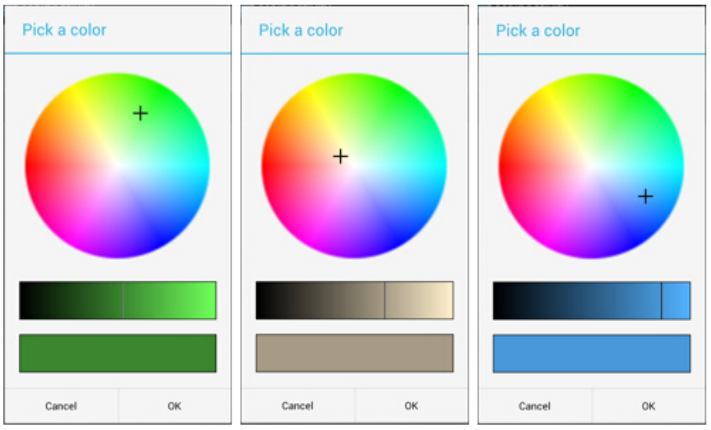


FIGURE 2.14: Picking a colour by using *V-HS* colour wheel.

Bar-Square The combination of *Bar-Square* is most popular at present. The linear slider could represent the *hue*, or the *saturation*, or the *value*, therefore there are three arrangements of a slider bar available: *hue* slider, *saturation*

slider, and *value* slider. For example, the *H-SV*, *S-HV*, and *V-HS* colour square shown in Figure 2.15.

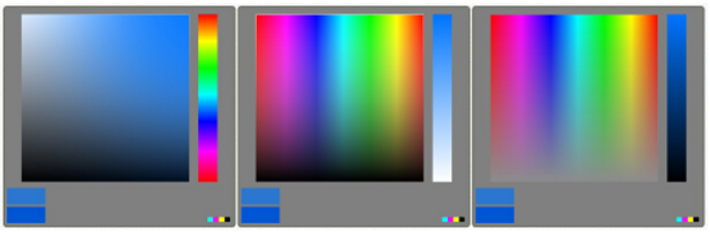


FIGURE 2.15: Representations of *H-SV*, *S-HV*, and *V-HS* colour square. (taken from https://community.kde.org/Krita/Community_Mockups_and_Wishlist)

Most often the *Microsoft*^{TM} (MS) *Windows* software adopts the *Bar-Square* combination since the release of *MS Office*^{TM}95. For example, the *L-HS* colour square is used in *MS Paint*^{TM} and *MS Word*^{TM}, see Figure 2.16. The colour picker adopted in *MS Word*^{TM} uses the RGB colour space by default, as presented in Figure 2.8 (right), and the pull-down menu allows for switching between the RGB and HSL colour space. The 2D *hue-saturation* square is statically coloured, and one auxiliary *lightness* slider on the left is dynamically coloured. The windows of the colour selector can be freely resized, with no aspect ratio constraint.

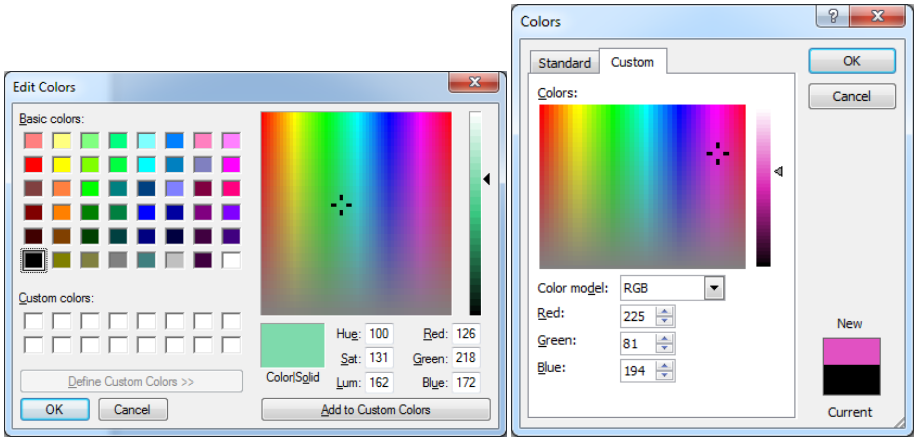


FIGURE 2.16: Screenshots of colour pickers in *MS Paint*^{TM} (left) and *MS Word*^{TM} 2010 (right).

The 2D input square could be divided equally into sets of small boxes. For instance, there is a pull-down menu that can switch *saturation-value* square type between the 'smooth' or the 'discrete' in *ColorZilla*[3] colour picker, as shown in Figure 2.17.

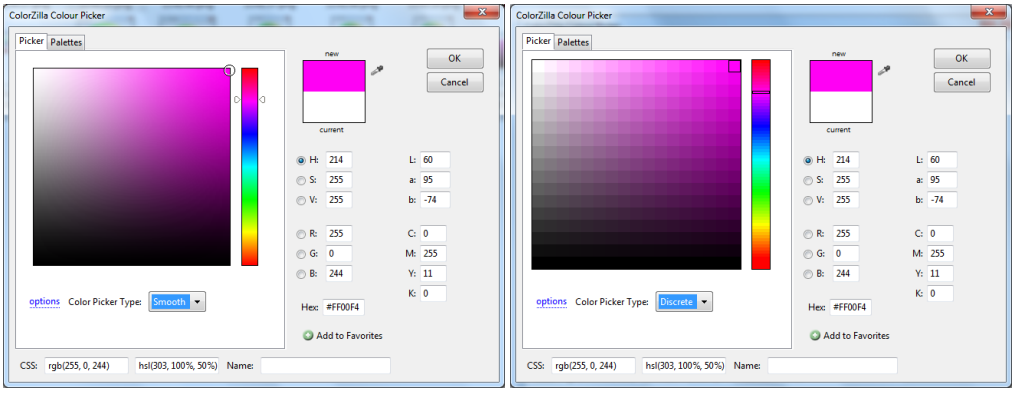


FIGURE 2.17: Screenshots of *ColorZilla* colour picker.

Ring-Square/Triangle In most cases, a *hue*-based colour picker is represented to the user in form of a linear or a circular *hue* slider and a two-dimensional area, which is usually a *square* or a *triangle*. A colour can be chosen by firstly picking a full-intensity *hue* value within the *colour ring*, then selecting the desired *saturation* and *lightness* from a square region (e.g., Figure 2.18, left) or a triangle region (e.g., Figure 2.18, right).

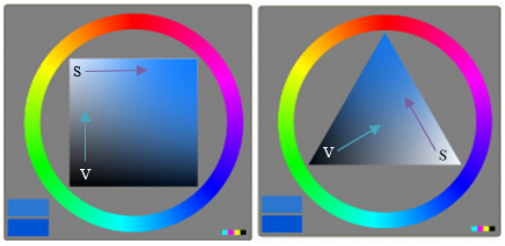


FIGURE 2.18: Representation of a H-SV *ring-square* (left) and H-SV *ring-triangle* (right). (taken from https://community.kde.org/Krita/Community_Mockups_and_Wishlist

[3] http://www.colorzilla.com/

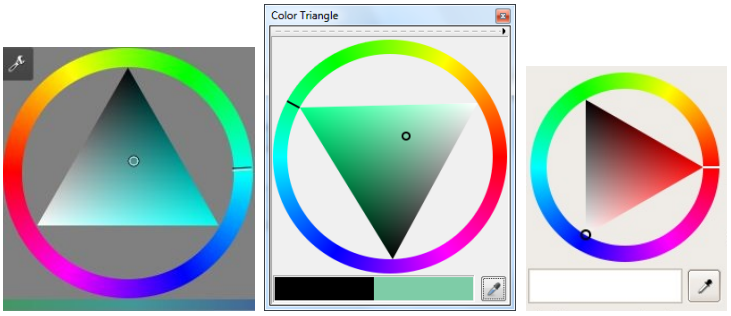


FIGURE 2.19: Screenshots of *Krita* colour picker (left), *My Paint Color Triangle* (middle), and *GTK+* colour picker (right).

The square or triangle region in a traditional *Ring-Square* or *Ring-Triangle* colour picker is static. Taking the triangle input area for example, the triangle area of the colour picker of *Krita*[4] always stays in the same direction (see Figure 2.19, left). While in some current applications, such as the *Color Triangle* in *My Paint*[5] (see Figure 2.19, middle) and *GTK+*[6] colour picker (see Figure 2.19, right), the inner triangle region is moved together with the selected *hue* on the colour wheel.

More than Two Shapes Many programs also allow the user to select a colour via multiple colour pickers with different shapes and some numeric boxes at the same time. The alteration of a visual colour selection tool will cause the corresponding changes of the other forms of colour picker, and vice versa. For example, the HSV colour picker in *Colourlovers*[7] shown in Figure 2.20, the *saturation-value* square, the *hue* slider, the *saturation* slider, the *value* slider, and the *Hexdecimal* (HEX) value, RGB value, HSV value and CMYK value boxes are shown together to the user. *ColorImpact*[8] (see Figure 2.21) includes a *hue-saturation* wheel, *hue* slider, *saturation* slider, *brightness* slider, *HEX* value boxes as well as a customised colour ring that segments a continuous colour

[4] https://krita.org/
[5] http://mypaint.intilinux.com/
[6] http://gcolor2.sourceforge.net/
[7] http://www.colourlovers.com/copaso/ColorPaletteSoftware
[8] http://www.tigercolor.com/



FIGURE 2.20: Screenshots of the HSV colour picker of *Colourlovers*.

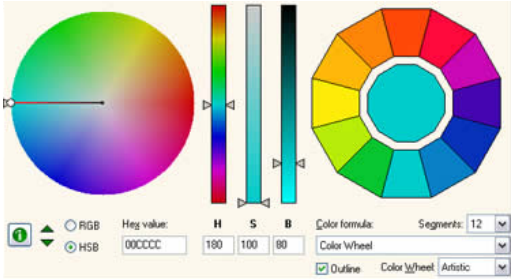


FIGURE 2.21: Screenshot of *ColorImpact* colour picker.

ring into a pre-defined number of colours. The changes of any of these elements would immediately interact with the other parts of a colour picker.

To give full freedom to the users, one direct solution is to make it possible for users to choose the colour selection methods that work best for them. One example is the *CoolPicker*[9] plugin for *3DS Max*, which integrates a *triangle-wheel* and a Mixer with the sliders and the numeric values of each dimension of RGB, HSV, CMYK, LIQ (*Luma-Inphase-Quad*) and *alpha* value, and 12 colour planes (RG, GB, RB, HS, SV, HV, CM, MY, CY, LI, IQ, and LQ square), see Figure 2.22.

[9] http://www.idigitalhouse.com/Tech/dh_tech.html

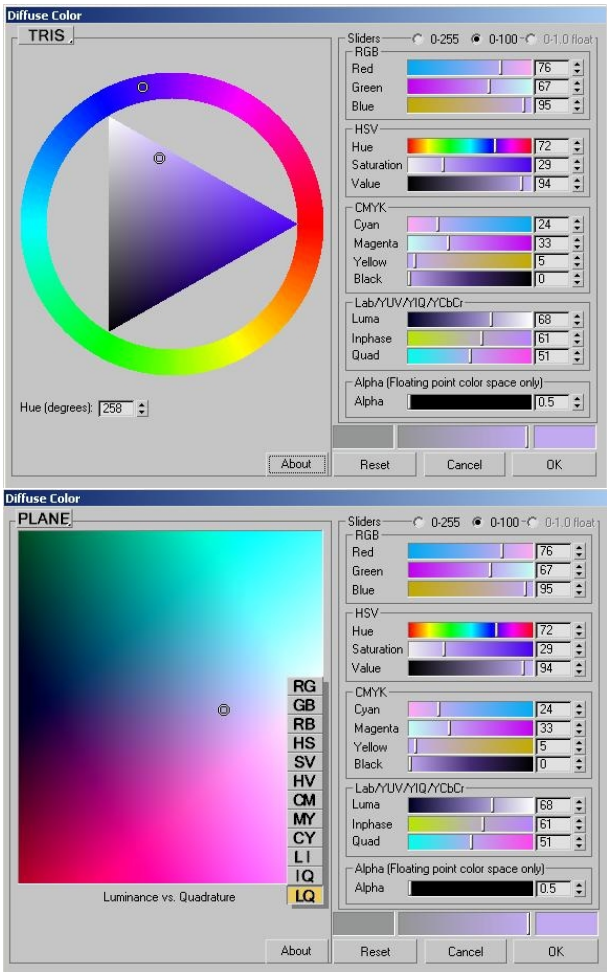


FIGURE 2.22: Screenshots of user interfaces of *CoolPicker* plugin for *3DS Max*.

2.2.2.3 Colour picker Categories

There are different ways of categorising colour pickers. Based on the number of colour choices, two categories of colour pickers are recognised: one is the discrete palette which contains a limited number of colours, and other is the continuous palette which has an infinite number of colour choices. Cynthia Brewer [13] further divided the discrete colour palette into three basic sub-categories for data visualisation in cartography:

- *Sequential palette* - displays one-hue based colours with different shades in an order changing gradually from less saturated to full saturated colour,

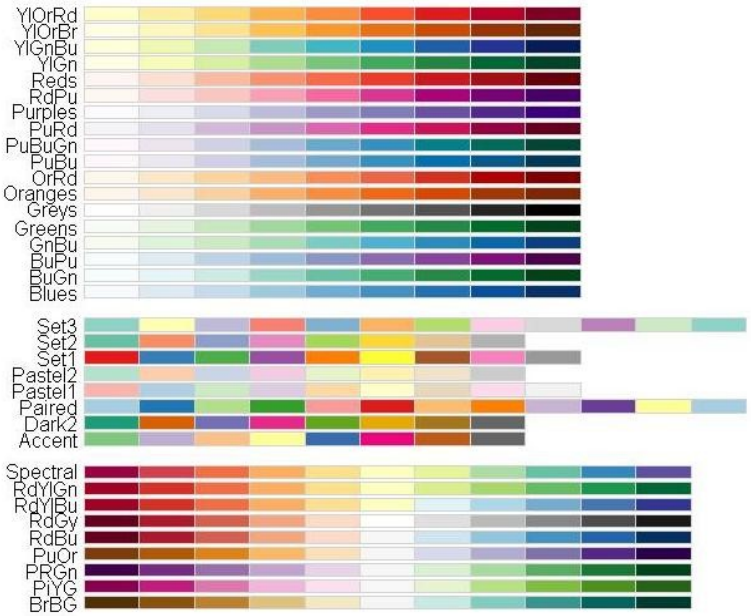


FIGURE 2.23: Brewer palettes in *RColorBrewer* package of R.

and the difference between successive colours is perceived to be uniform.

- *Qualitative palette* - displays colours of different hue values with equal *lightness* and *saturation*.

- *Diverging palette* - a concatenation of two *sequential* palettes with different hue values.

Brewer hand-picked a collection of colours and fabricated a colour palette known as Brewer Palette [14]. Figure 2.23 shows the Brewer Palette in a colour package '*RColorBrewer*' of R package[10]. The top of the groups are 18 *sequential* palettes which display in a clear luminance order with equal monotonic distances; the middle groups are 8 *qualitative* palettes created by varying hues in maximum way of distinguishable colours; 9 *diverging* palettes are illustrated by the bottom groups, which have two dark colours at the end-point and one light colour at the mid-point.

[10] http://www.stat.ubc.ca/~jenny/STAT545A/block14_colors.html

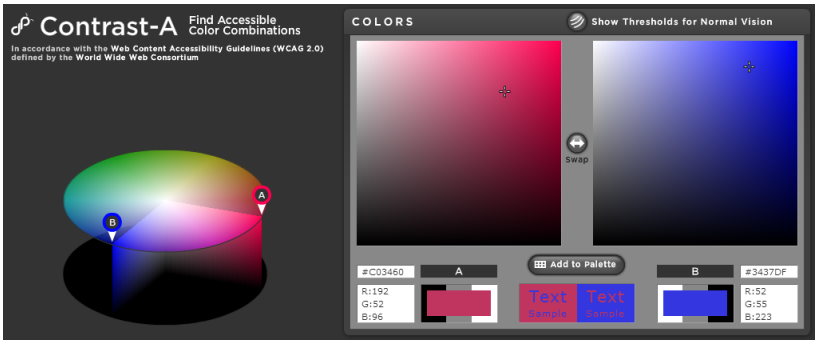


FIGURE 2.24: Screenshot of 3D colour picker of *ContrastA*.

Colour pickers could also be grouped into three categories according to how the three dimensions of a colour are manipulated:

- The '*1D-1D-1D*' colour picker - allows only one of the three parameters to change at a time.
- The '*1D-2D*' colour picker - contains a two-dimensional input area which allows for changing two parameters simultaneously, and a slider that allows access to the third parameter.
- The '*3D*' colour picker a slice through a 3D sphere is firstly cut, followed by selecting a colour in a 2D plane. Three-dimensional input techniques are not yet widely used [112].

3D Colour Picker Wu and Takatsuka [149] designed a three-dimensional (3D) RGB colour picker, a 3D HSV colour picker and a 3D Munsell colour picker, and proposed that 3D colour palettes would facilitate users' colour picking. We only found one real-life example of a 3D colour picker, *ContrastA*[11] (as shown in Figure 2.24), which is used for designing colours for the font and background of web pages. The users can first select a colour in the top colour circle on the left-hand side, and the *saturation-value* planes on the right-hand

[11] http://www.dasplankton.de/ContrastA/

side would change accordingly, and the final colour is chosen from the colour square.

2.2.3 Existing Colour Pickers

2.2.3.1 Colour Pickers used in CBIR

User interfaces represent the visible part of Content-based Image Retrieval (CBIR) systems. Among many present query paradigms, only the *Query-by-Colour* and *Query-by-Drawing* methods require the user interface of colour selection, from which users need to pick colour(s) as the direct colour query input or as pen or brush colour to draw a query input.

Colour Picker used in Query-by-Colour System For colour-based CBIR image search systems, colour plays an important role in the accuracy of search algorithm. Due to the lack of colour picker interface research documented in current Query-by-Colour CBIR literature, we also reviewed several commercial web-based CBIR systems which mainly use colour features for search. There are two kinds of colour pickers adopted: the discrete colour palette with a limited number of colours and the continuous colour picker. A typical example of adopting a discrete colour palette is *Google Images*[12], as shown in Figure 2.25, where users only need to select a pre-defined colour to narrow down their search results. The *hue* slider is used by *Shutterstock Spectrum*[13] but with a limited number of colour choices as shown in Figure 2.26. Both *Shutterstock Palette*[14] (see Figure 2.27) and *MultiColr*[15] (see Figure 2.28) allow users to select more than one key colour (up to 5 colours) from a medium number of

[12] https://images.google.com/
[13] http://www.shutterstock.com/labs/spectrum/
[14] http://www.shutterstock.com/labs/palette/?image_type=photo&keyword=
[15] http://labs.tineye.com/multicolr/

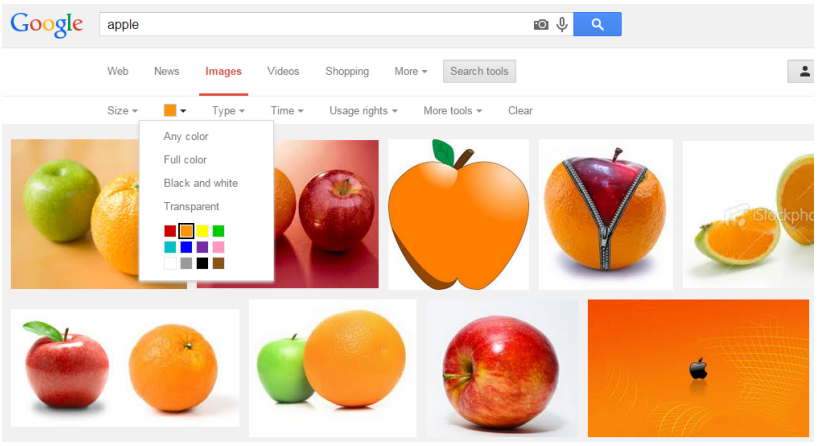


FIGURE 2.25: The colour palettes used to filter keyword-based search results on *Google Images*.

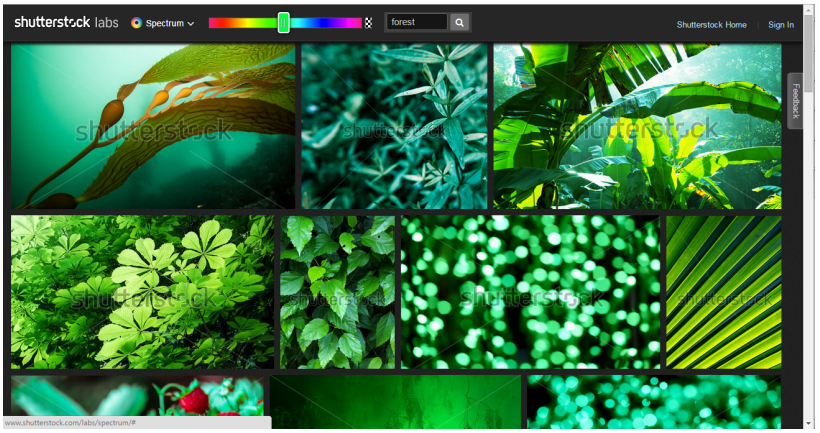


FIGURE 2.26: Screenshot of *Shutterstock Spectrum*.

gradient colours to search for the images containing the selected colours. The differences between these two systems lie in the number of colour choices and the colour values.

Only a few web image search engines employ continuous colour pickers to give users freedom to mix any colour by themselves. For example, *Xcavator.net* (as shown in Figure 2.29) enhances a keyword-based search by allowing users to select a specific colour via a H-SV colour wheel or to upload an exemplary image as a query input.

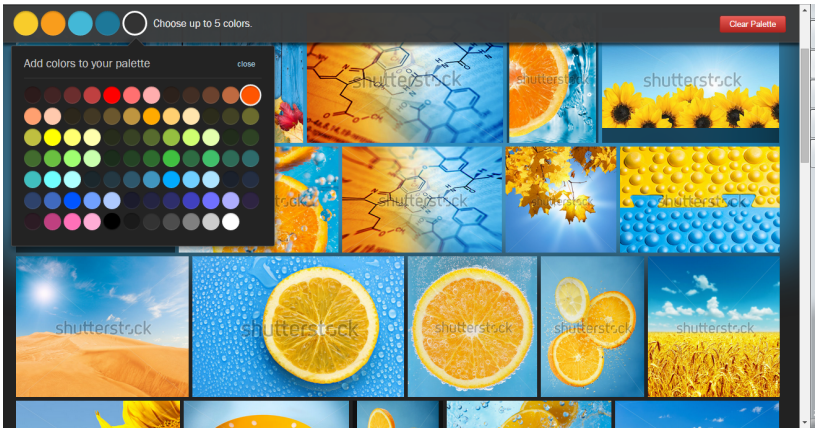


FIGURE 2.27: Screenshot of *Shutterstock Palette*.



FIGURE 2.28: Screenshot of *Multicolr* from *TinEye Labs*.

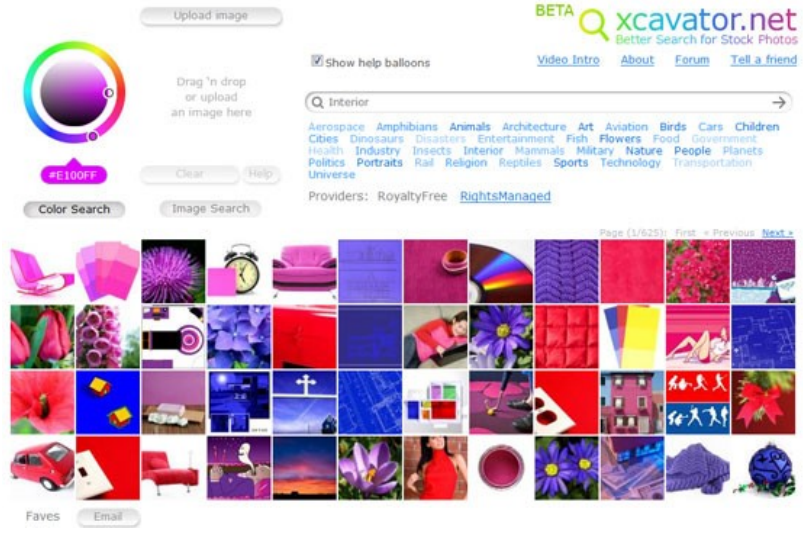


FIGURE 2.29: Screenshot of *xcavator.net*.

Colour Pickers used in Query-by-Drawing System Three categories of colour pickers used in drawing-based CBIR image search system are recognised: the *arbitrary discrete* colour picker, the *gradient discrete* colour picker and the *continuous* colour picker.

The arbitrary discrete colour picker is subjectively designed by system developers. For example, *ColorMap* [142] allows users to select a colour from 12 focal colour swatches to generate or refine a colour layout in a composite canvas as the visual query input (see Figure 2.30, left); similarly, the *PICASSO* [27] system offers 20 colour choices for users to compose a simple colour layout to search for paintings, as shown in Figure 2.30 (middle); and the *VISI* system [65] provides 40 specific colours for users to pick a colour for the pen and shapes (see Figure 2.30, right). These researchers did not explain why they chose these colours, neither on which rules they positioned these colours, nor what the values of these colours were.

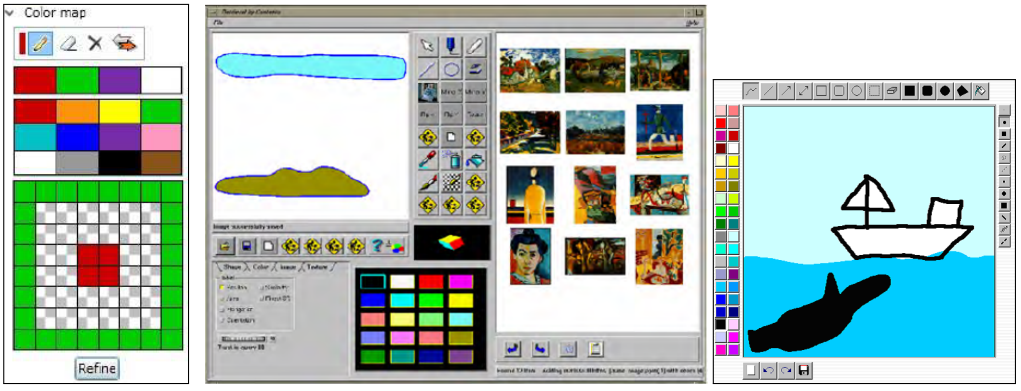


FIGURE 2.30: Screenshots of the query canvas of *ColorMap* (left), *PICASSO* (middle) system, and *VISI* system (right).

Gradient discrete colour pickers are used more often than discrete colour picker with arbitrary colours. For example, the *Search Google by Drawing*[16] website offers users colour swatches with linear gradient colours and brushes with 14 different sizes to draw a query and then submit the drawing to the *Google Images* search engine. As can be seen in Figure 2.31, the search results are not satisfied.

[16] http://search-by-drawing.fullstackoptimization.com/

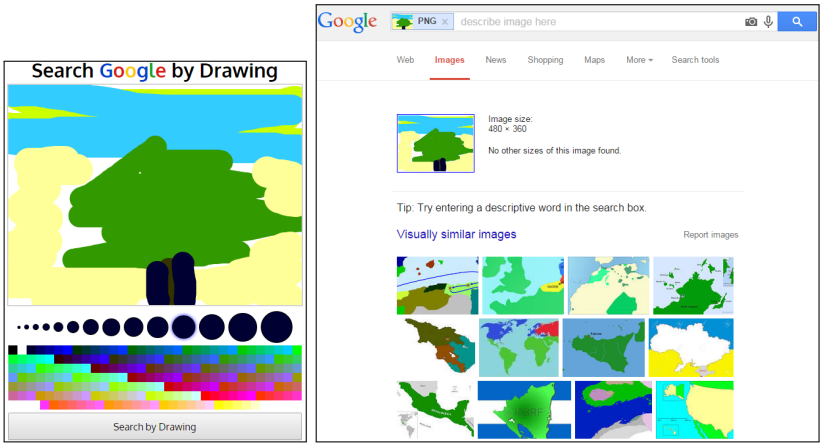


FIGURE 2.31: Screenshots of *Search Google by Drawing* (left) and search results of a tree query (right).

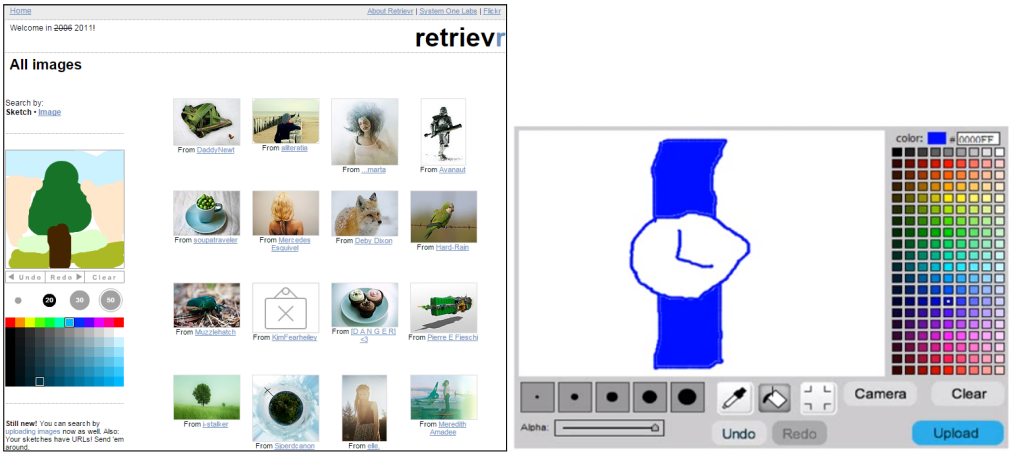


FIGURE 2.32: Screenshots of the user interfaces of *Retrievr* (left) and *GazoPa* (right).

Similar online image search systems are *Retrievr*[17] and *GazoPa*[18] (see Figure 2.32), which offer a limited number of colour choices and a small number of pen sizes.

Regarding continuous colour pickers, different researchers adopted different shapes and colour models for a colour selection tool. It is found that the HSV colour model is the most frequently used colour model by Query-by-Drawing

[17] http://labs.systemone.at/retrievr/
[18] http://www.gazopa.com/

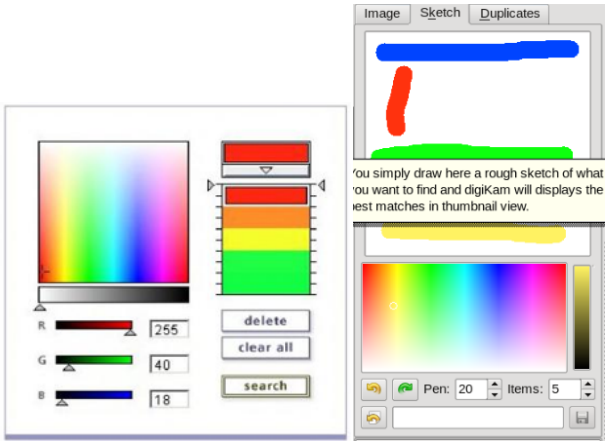


FIGURE 2.33: Screenshots of the colour picker used in *QBIC* (left) and *digiKam* (right).

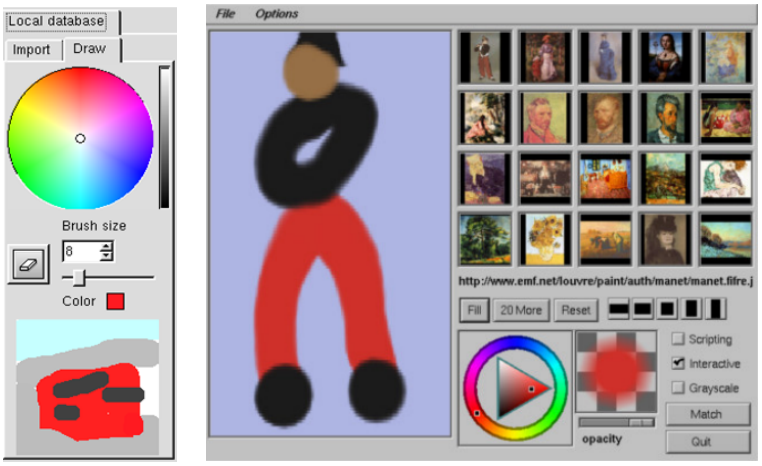


FIGURE 2.34: Screenshot of the colour picker and canvas used in *imgSeek* (left) and Jacobs et al.'s (1995) [69] system (right).

CBIR image search systems. For example, both IBM's *Query By Image Content (QBIC)* [45] and *digiKam*[19] adopt a V-HS colour square, as shown in Figure 2.33. Similarly, based on HSV colour space, *imgSeek*[20] uses a colour wheel, and a *ring-triangle* colour picker is employed in [69], see Figure 2.34.

[19] https://www.digikam.org/
[20] http://www.imgseek.net/

2.2.3.2 Colour Pickers used in Current Mobile Applications

The intuitive nature of mobility and embedded touchscreens of mobile devices allow artists or non-artists to draw freely. We installed and reviewed published mobile applications (Apps) by undertaking searches with keywords and phrases related to '*draw*', '*drawing*', '*sketch*', '*photo editor*', '*colouring*', and '*note taking*' on *Google Play store*[21]. By the time of writing, we have reviewed 140 drawing Apps, 24 colouring Apps, 18 photo-editor Apps, and 7 note-taking Apps, which are categorised based on functional purpose.

TABLE 2.1: Distributions of colour pickers used in reviewed Android Apps on *GooglePlay*.

Apps for	Discrete colour picker	Continuous colour picker	With one colour picker	With two colour pickers	With three colour pickers	BG-colour picker
Drawing (140)	69	114	93	30	17	42
Colouring (24)	22	12	13	10	1	1
Photo-editor (18)	4	16	12	45	1	0
Note-taking (7)	4	4	5	1	1	1

It is found that a colour selection task is required for various purposes, from picking an approximate colour to making notes to obtaining a colour with a specific shade, e.g., to draw a banana. As shown in Table 2.1, 91.67% of the reviewed 'colouring' Apps adopt a discrete colour palette (although the number of colours differs); 81.43% of the 'drawing' Apps and 88.89% of the 'photo-editing' Apps provide a continuous colour picker. It is noted that some Apps contain two or three or even more than three colour pickers, as shown in Figure 2.35. 33.57% of the drawing Apps have at least two colour pickers. For example, the drawing App *ArtFlow* includes two continuous colour pickers (H-SV colour *ring-triangle* and RGB/HSV sliders) and three discrete colour palettes, see two screenshots on the right-hand side in Figure 2.35.

[21] https://play.google.com

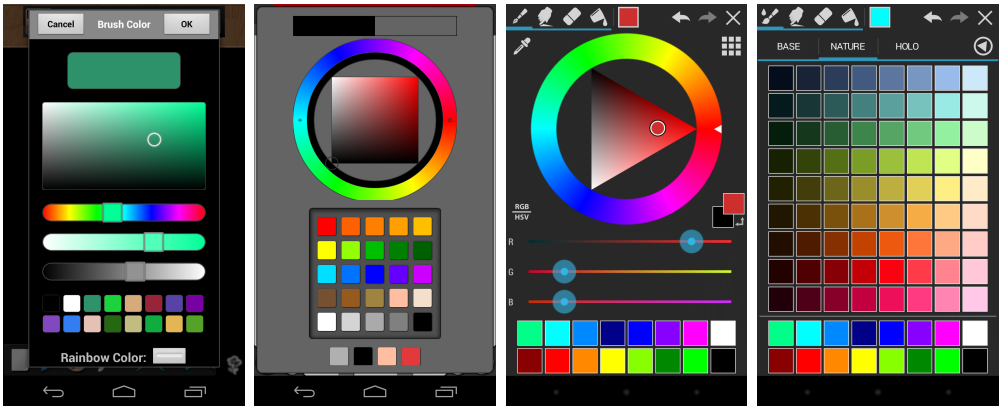


FIGURE 2.35: Screenshots of Apps that use multiple colour pickers. Left-to-right: *PaintJoy*, *InfiniteDesign*, *ArtFlow*-continuous picker and *ArtFlow*-discrete picker.

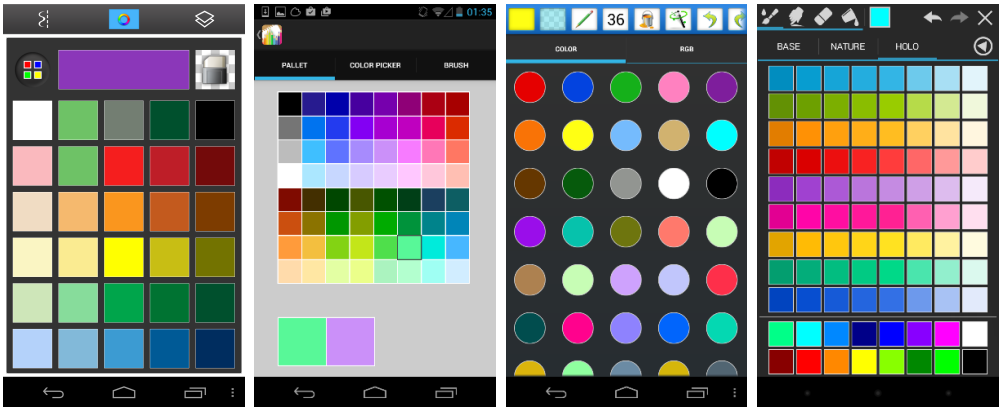


FIGURE 2.36: Screenshots of the discrete colour palette used by *SketchBookX*, *Spline*, *PaintFree*, and *ArtFlow*.

For continuous colour pickers, there are no commonly accepted guidelines on designing an 'appropriate' shape of the colour representation. The *square*, *triangle*, *ring*, *wheel*, and *bar* shape are commonly recognised from the reviewed colour pickers. The most commonly used shape is the *square*, followed by the *bar*, the *ring* and the *wheel*.

The application developers choose different discrete colour palettes for different applications. The interface elements of these discrete colour palettes, such as the shape of the colour swatch (*circle* or *square*), the number of colours, and the values of colours, differ significantly (as shown in Figure 2.36). One of the restrictions of discrete colour pickers is the limited number of colour choices.

Increasing pre-defined colours is one solution commonly adopted. An illustrative example comes from the drawing application *MagicDoodle* which contains 16 groups of colour palettes, such as a '*pencil color*' palette, a '*skin tone*' palette, and a '*van Gogh oil*' palette, as shown in Figure 2.37.

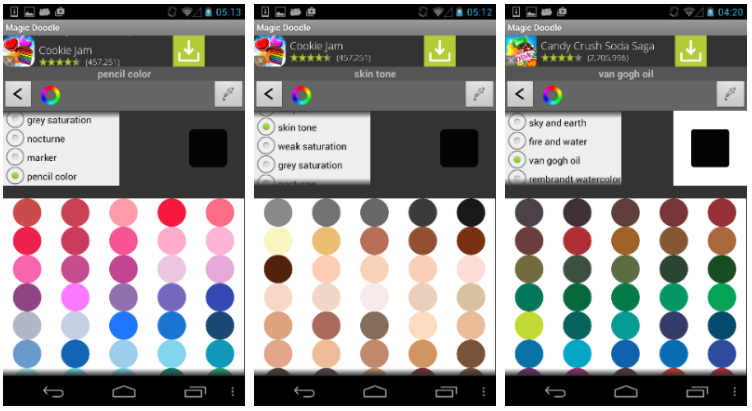


FIGURE 2.37: Screenshots of three discrete colour palettes out of 16 colour palettes used in *MagicDoodle*.

Another intuitive solution is adopting a compromise between continuous colour pickers and the discrete ones. In other words, when a colour is chosen, a limited set of similar colours with different levels of *saturation* or *brightness* are automatically generated. For example, as shown in Figure 2.38, *PixlrExpress* supplies 15 shades of a chosen colour in a fully-saturated *hue* slider; and *WatercolorPencil Lite* provides 32 shades of a selected colour in a colour circle.

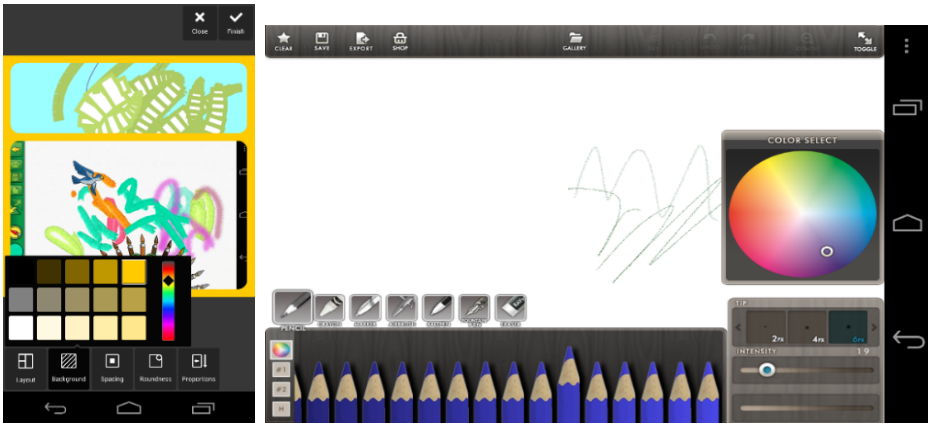


FIGURE 2.38: Screenshots of *PixlrExpress* (15 shades of one hue), *WatercolorPencil Lite* (32 shades of one hue).

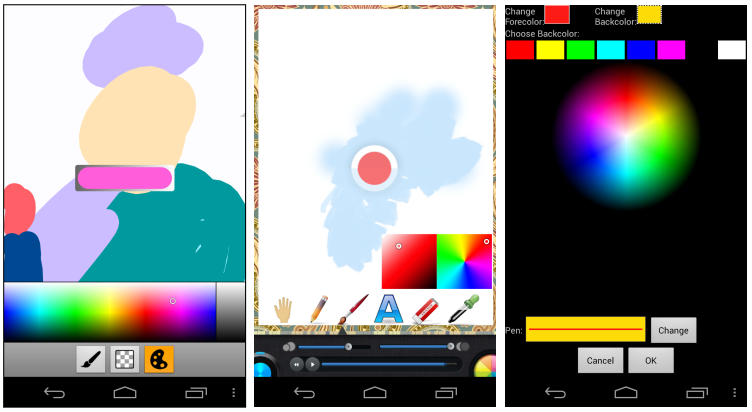


FIGURE 2.39: Screenshot of the colour pickers that consider colour context. Left-to-right: *FingerColors*, *AirBrush*, and *LomoPaint*.

Most colour pickers enable the user to select colours without considering the use of colours in the context. Typically, the colour picker is implemented as a form of colour selection dialogue, and the selected colour is shown on a small swatch rather than on the context where the colour would be applied. It was mentioned by Moretti [96] that a colour chosen out of context almost looks different when being applied to the interface. There are a few exceptions of applications such as *FingerColors* and *AirBrush* (see Figure 2.39), which display the selected colour on a small region on top of the canvas. However, the input area of colour picker is too small for the fat finger. In its colour selector dialogue, *LomoPaint* offers a swatch that dynamically shows the modified colour (line) together with the current background colour, as shown in Figure 2.39 (right).

2.2.4 A Survey on Mobile Phone Use and Drawing Frequency

We have conducted an online survey via *Survey Monkey* to explore the types of mobile phones owned by the general public and how frequently people draw, as well as the painting type people prefer to view (see Section 4.3.1.1). In total, 134 responses (88 male and 45 female) were collected with 115 responses on the drawing frequency. The results suggest that most of the responders (93.28% of 134 responses) use high-end touchscreen mobile phone, and 84.33% of these

smartphones have the capability of internet connection. 29.10% indicate they have experience of drawing on touchscreen phones. However, 46.09% of 115 responders do not draw at all, see Figure 2.40.

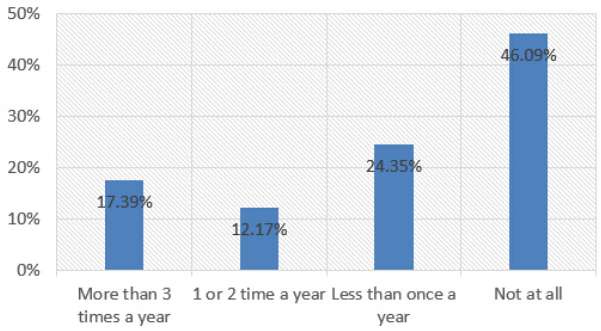


FIGURE 2.40: A survey result of 115 responses on the frequency of drawing.

2.3 Related Work of Colour Selection Study

As mentioned in the previous section, a large number of colour picker interfaces have been proposed for both mobile applications and computer software and websites over the years. However, not much research has been conducted in the field of user interfaces of colour selection tools. Some researchers focussed on the discussion of which colour space or colour model is best for a specific purpose (please refer to Section 2.2.1). Colour matching experiments are often conducted for previous research on colour selection.

2.3.1 Colour Matching Experiment

Colour matching tasks have been used for colorimetry research and psychological study for a long period. During earlier colour matching experiments [150], the participant is asked to mix two or more standard coloured lights by presenting them on a black background until the target coloured light is produced. Afterwards Perez-Carpinell et al. [106] carried out a study to ask participants to

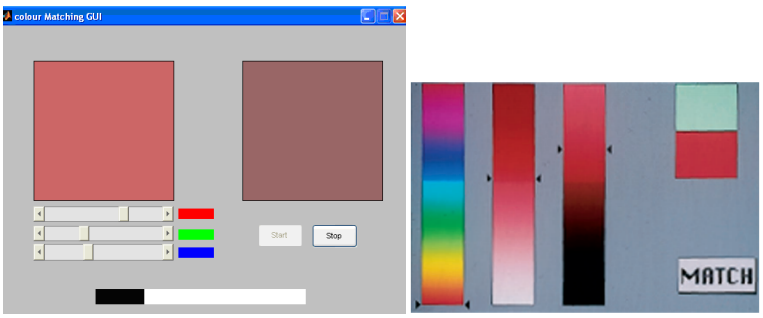


FIGURE 2.41: The setups used for colour matching in Henry et al. [64] and Douglas & Kirkpatrick [39].

select a physical (printed) Munsell colour chip that best matches with the one seen before. Modern colour matching tasks are conducted on electronic displays with wide colour gamut such as CRT (cathode-ray tube) screens [133]. The design for current colour matching experiments often contains a simple user interface for displaying a target colour and the participant's selected colour together, as well as a colour selection tool. Participants adjust a controlled colour until it matches the target colour. For instance, the setup of Henry et al.'s [64] time-limited colour matching experiment (shown in Figure 2.41, left) adopts RGB sliders for manipulating colour, and a bottom bar indicates the time remaining for a task. Some researchers did not restrict the time per participant to complete one match such as Douglas and Kirkpatrick [39]. For strict psychological experiments, the distance between the participant and the display screen is fixed as well.

A colour matching task can be thought as a component of a real-world colour selection task, because at least the physical actions of a user interacting with a colour picker are the same for both types of tasks. Hence, some researchers adopt the colour matching task as a simplification of a real-world colour specification task. However, using a colour matching task as a representative of real task has both advantages and limitations. Because all colour selection tasks

involve the same process of navigating through a colour space, the colour matching task offers a capability of an objective analysis of colour selection performance, such as the matching time and the accuracy of the selected colour. Thus, different systems can be quantitatively compared based on these measures. The disadvantages lie in the differences between a colour matching task and a real-world colour specification task. First, there is no specific motivation of matching colours nor the complex contexts around a selected colour. Second, the target colour is visually presented in close proximity in the screen of a colour matching task, while a typical colour specification task requires the mental colour in the user's mind [145]. Third, lab-based colour matching experiments are conducted under strictly-controlled display and viewing conditions. Therefore, observing the colour specification in a real use context of computer or mobile application is necessary in this regard.

Lots of researchers have found the colour shift of colour matching tasks across media [58, 73]. Therefore, in our experiments we decide to use a drawing tool installed on one touchscreen phone with the same settings (e.g., the level of brightness) as those of other phone displaying reference paintings.

2.3.2 Visual Design of Colour Picker

In this thesis, the visual design of a colour picker refers to everything related to the user interaction with a colour picker, for example, the dimension, size, orientation, shape and rendering style of the input area that allows the user to freely manipulate colours. Until now, to the best of our knowledge, there are no published academic reviews on the visual design of colour selection user interface nor design guidelines documented in the graphics industry. Some of the methods for previous study design and measures are useful for our experimental design.

According to Wells [145], a colour picker contains five general attributes, which are *colour model*, *input method*, *drawing style*, *context*, and *visual design*. Wells [145] recognised three types of input area rendering style - *plain*, *static*, and *dynamic* (only 1D sliders are used in her work), and two kinds of background context of a colour being modified - *achromatic* and *chromatic*. The plain slider is always black-and-white; the static slider is statically coloured with a full range of allowable values and it never changes during colour selection; the dynamic slider is dynamically coloured as the parameters of selected colour being manipulated. With the combinations of three sliders, three colour models, and two contexts, Wells [145] explored 18 colour pickers in a colour matching experiment, in which 83 undergraduate students participated. Wells collected and analysed the following measures: the *time of selecting a colour*, the accuracy (including *overall colour*, *hue*, *chroma*, and *intensity error*), the physical effort (*slider access* and *slider time*) and subjective preference. The results show that the participants generally preferred the dynamic slider over the static slider and the static slider over the plain slider, with the exception of using RGB colour model where the static slider was preferred over the plain one. The colour matching performance of using the dynamic slider was superior to that of using the static slider followed by using the plain slider. The achromatic background behind the controlled colour swatch led to more accurate colour than the chromatic context.

These results are consistent with the conclusions made by Douglas and Kirkpatrick [39], who studied sliders with and without feedback by asking 48 undergraduates to perform colour matching tasks on four colour pickers:*RGB-no feedback*, *RGB-feedback*, *HSV-no feedback*, and *HSV-feedback* (as shown in Figure 2.42). They concluded that visual feedback was the most important factor that influences human performance, such as matching time and accuracy. Douglas and Kirkpatrick [38, 39] only studied the '*1D-1D-1D*' control of colours, but also discussed '*2D*' and '*3D*' controls, and proposed a hypothesis that colour

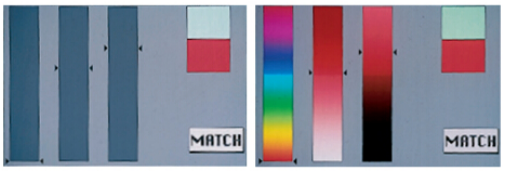


FIGURE 2.42: The sliders with feedback (left) and without feedback (right) used in Douglas&Kirkpatrick's study [39].



FIGURE 2.43: Five colour selector interfaces used in Kok's study [76].

selection performance with '*3D*' control might be the best among the three. However, this proposal has not been evaluated empirically.

Kok [76] proposed a *Scientific User Interface Testing* framework (SUIT) to evaluate the usability of five colour selection interfaces including a colour palette with 30 colours, *Microsoft Office*^{TM} custom colour picker, 68 basic colours of *Microsoft Paint*^{TM}, *Microsoft Office*^{TM} standard colour picker (144 colours), and *Vind(x)* colour picker [139] which enables users to mix two of 11 colour categories, see Figure 2.43. Having restricted time and resources, each participant was asked to reproduce the previously-viewed colours on the computer. The *mouse activity*, the *selected colour accuracy*, *reaction time* as well as *subjective preference* were collected and listed without further analysed. Kok proposed that neither handedness nor computer experience of participants affected the accuracy of colour selection. The participants with some colour experience chose colours more accurately and slower than those without any colour experience. Finally, Kok discussed each colour picker in four aspects [76]: colour availability (the number of colour choices), intended user group, learnability, and the method of selection (sliders or clickable areas). He concluded that there was no best colour selector, and some colour pickers produced a higher selection accuracy but it took a longer time to select a colour.

The capability and effectiveness of multi-touch colour selectors were explored in [51]. The authors designed a tablet application with a multi-touch V-HS colour wheel, named *ColorFingers*, as shown in Figure 2.44. Users have to select a colour from a colour wheel firstly by using one finger and then moving another finger up or down to change the *lightness*, and the selected colour is dynamically displayed on a colour ring around the hue-saturation colour wheel. With the target colour shown on the left and right edges of the screen (see Figure 2.44, right). A between-subjects colour matching experiment involving 20 university students was conducted on the *ColorFingers* and Douglas and Kirkpatrick's experiment settings [39]. The authors claimed that the performance (both speed and accuracy) of the colour matching using *ColorFingers* was better than Douglas and Kirkpatrick's sliders [39].

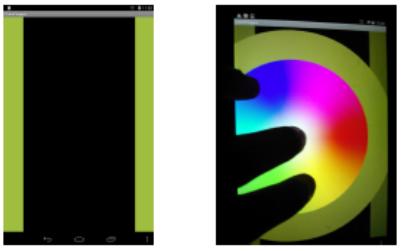


FIGURE 2.44: Screenshots with target colour (left) and colour matching process on *ColorFingers* (right) (taken from [51]).

2.3.3 Colour Selection Study in CBIR

To the best of our knowledge, little work related to colour selection interfaces of the content-based image retrieval (CBIR) systems exists in the literature. Van den Broek et al. [134, 135] reported that they were the first to conduct a review of the colour selection interfaces used in CBIR and categorised ten colour selection tools into three groups: slider-bars, discrete colour matrices, and continuous colour squares. However, actually they only reviewed three CBIR colour pickers (including *QBIC* [45], *SIMPLIcity*[22], *VisualSEEK* [122]) and the others are

[22] http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi

taken from *PowerPoint*, *EditPlus* and three online Java Widgets. Van den Broek et al. [134] proposed that a limited number of colours are sufficient for a CBIR system, and existing colour selector interfaces are complex and provide unnecessary variations, such as the numeric input box and 2D colour square. They developed a colour picker *Vind(x)* to draw a query for a CBIR system [139], as shown in Figure 2.45. Only 11 focal colours and a slider bar are shown to the user and the *Vind(x)* colour selector allows the user to select a colour by mixing any two of the 11 colours.

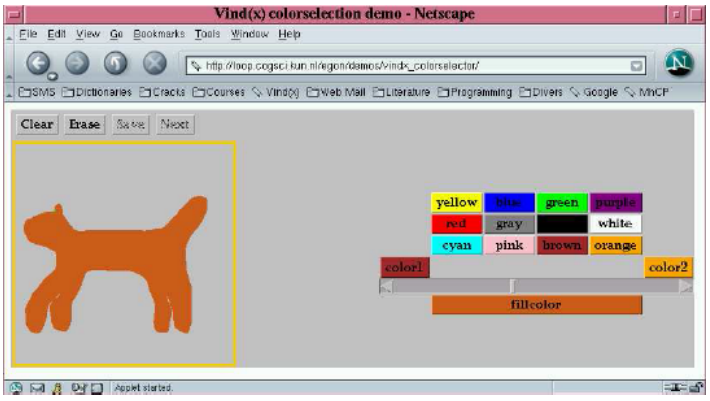


FIGURE 2.45: The user interfaces of *Vind(x)* system (taken from [139]).

McDonald and Tait [92] compared the performance of a sketch tool and a structured navigation tool of their system under different types of tasks. 24 undergraduates were asked to draw queries for seven target specific tasks and seven target non-specific (theme) tasks. The user satisfaction, the number of successful searches, and the time of completing a task were measured and analysed. They found that there was no significant difference between the use of sketch and browsing tools for those two types of tasks, except that the participants completed the target non-specific task by browsing the database more quickly than sketching a query. They also compared user satisfaction of their sketch tool with IBM's QBIC [45], and claimed that the participants preferred their sketch tool because of its design of freehand drawing tool.

Hove (2007) studied users' drawings on two web-based Query-by-Drawing image retrieval systems *VISI* and *Retrievr* (see Section 2.2.3.1). A new framework was proposed to analyse and evaluate participants' drawings to explore how the participants draw by using the freehand drawing tools and the pre-defined shapes. However, the limitations of this work lie in: 1) The two examined drawing systems are web-based, which means less intuitive interaction than drawing on touchscreen mobile devices; 2) Lack of connection between the user's drawings and the system's search results, which makes the study focussing on the participant drawing analysis out of the context of image retrieval system; 3) Hove's framework aims to categorise the users' drawings only according to the degree to which a drawing represents a realistic depiction of the objects in the drawing, while the Query-by-Drawing paradigm does not require a realistic query drawing; 4) The two web-based systems (both developed by someone else) adopt the same type of colour palette, so the study does not compare different kinds of colour picker designs; 5) Low number of participants - only 17 participants performed the drawing tasks in a specific domain (e.g., 'dolphin' in a zoo).

It is clear that the previous work was conducted on computer-based CBIR systems developed almost 20 years ago. The limitations include: 1) non-intuitive interaction of colour selection by using a mouse on the computer; 2) out-of-date colour palette interfaces; and 3) cannot be representative of the capability of current mobile device.

2.3.4 Colour Selection Strategy

As a common user interaction activity, colour selection has a variety of users, who may or may not have knowledge about how the colour model behind a colour picker works. According to Lyons et al. [90], most users are not familiar with the grounding principles of a colour space or the professional colour terms

and concepts [121]. Bauersfel et al. [8] suggested that a better colour picker would make itself easily accessible to a broader user population. It is proposed that understanding human behaviour is extremely important to make systems more effective and user-friendly [21]. To design an intuitive and understandable colour selection tool, it is believed that focussing on the user's selection strategy may be helpful and valuable. Wells and Tassinary [144] developed the *Standardized Performance Trajectory* (SPT) measure to analyse participants' colour matching behaviour. Bailey [5] claimed that the physical manner of selecting colours was significantly affected by the visual form of a colour picker.

The '*hill-climbing*' strategy was proposed and verified in [4]. Douglas and Kirkpatrick [38] found that a hill-climbing strategy was used quite often by non-expert users during colour selection process. That is, if moving in some direction which gets the colour closer, users will continue to move to that direction further; otherwise, users will backtrack and pick up colour in another direction. Douglas and Kirkpatrick [38] defined and separated out each move and calculated the colour difference between the resulting colour and target colour (measured in LAB colour space). By analysing the percentage of closer matches, they found that only a little more than half of all moves are towards the right direction. And the colour selection interface with visual feedback improved this accuracy. Therefore, they inferred that participants predicted the colour of next move by the interactive display of a colour picker [38], rather than reasoning based on the colour model.

Based on the same experiment settings, Douglas and Kirkpatrick [39] defined and analysed the *slider traces* of participants' each trial actions. They proposed that the characteristics of *slider traces* could reveal the participants' navigation behaviour of using colour selectors, such as the moving speed and the moving sequence of sliders, etc. Six slider-move patterns were recognised [39]: scanning, consecutive moves, goal regression, target crossing, pause, and jumping.

After analysing the moves for a single colour from four participants and comparing among each participant's performance, Douglas et al. suggested that the feedback help the participants with a certain level of knowledge of colour model to improve the accuracy of colour selection. However, they did not give a clear definition of the user with '*too little knowledge*', '*enough knowledge*' and '*much knowledge*'.

Douglas and Kirkpatrick [39] also proposed that users may adopt different search strategies when using colour pickers with different colour models. The *Maximal Repeating Pattern* (MRP) algorithm was proposed by Siochi and Ehrich [119] to recognise usability issues of interfaces, and the participants' actions were categorised into several states. With MRP analysis, the frequently occurring state transition patterns are considered as representative use actions (strategies). Goodrum et al. [56] are the first to adopt MRP analysis to explore the sequence of search moves made during image search on the web. However, little attention has been paid on user behaviour in *Query-by-Drawing* (QbD) image search, even less on the use pattern on QbD image search mobile applications. It is expected to adopt these previous methods to analyse the user behaviour of using our QbD mobile application in this PhD project.

2.4 Drawing-Related work

Humans have used drawing as a communication tool since the early cave age. Ancient cave paintings are found on the walls of caves in Europe since 40,000 B.C. [71]. Drawing plays important roles in many areas of human activities, such as note-taking, fashion, indoor design and other visual design related domains. With the development of technology and digital devices, artists begin to create artworks with the aid of computers. Increasingly high-processing capabilities and recent interactive technologies and touch-based interfaces provide

new possibilities for both artists and general users to draw digitally. Digital drawing offers several advantages, e.g., undo or redo of drawn strokes, never run out of paint, no dry time, etc. Colour drawing is a superior way to represent ideas and concepts in people's mind. Drawings can convey information that is difficult to express by words. According to Mohammad and Nishida [95], a very rough idea in a user's mind can be transferred into a concrete drawing through a repeating drawing-refining process. We aim to gain some methodologies and the best practices for our experimental design and result analysis by reviewing the literature of drawing-related work.

2.4.1 Psychological Studies on Drawing

Psychologists have studied the factors that affect people's drawing ability and quality of drawing, such as visual perception and cognition, etc. Cohen and Bennett [25] analysed drawing processes and concluded that the perception of the to-be-drawn stimulus was the most important factor. Drawing quality was less affected by shape constancy [24] and invasion [26].

The psychological drawing study in history usually conducts an experiment to ask a few artists or novices to make line drawings on paper by tracing the reference colour photographs under certain conditions - e.g., directly tracing the reference photos or rendering by looking at the photos [25]. In a drawing task, participants were allowed 10 minutes to realistically render the objects in the colour photograph [24]. During the analysis phase, researchers usually recruited human evaluators to rate drawings according to different properties of a drawing such as accuracy [26], aesthetic property [79], or overall quality [94]. Carson [18] proposed two objective measures of drawing accuracy (angel-based measure and polygon-based measure) and validated them with both expert and novice evaluators grading each drawing on a scale from 1 ('*poor*') to 10 ('*excellent*'). Cohen [23] suggested to find people without special knowledge of

drawing to be evaluators, while Ostrofsky et al. [103] came to an opposite opinion, in that only artists could be asked to be evaluators. Cohen and Earls [24] found that both the novices and experts produced almost similar ratings of portrait drawing. Kozbelt and Serafin [79] found that non-artists tended to relate quality with accuracy while art professors equated quality with originality.

2.4.2 Drawing as Query Input

Many current image retrieval systems use a drawing as a query input. The priorities of a drawn-query than other query forms lie in the capability of allowing users to clearly specify the information about colour, texture, proportion, layout, spatial relations as well as simple semantic meaning. Del Bimbo [34] argued that using *Query-by-Drawing* (QbD) tools is the primary means of submitting visual queries directly to CBIR system. As Banfi [6] concluded, there are many different forms of drawn query, including drawing with object contour, drawing with pre-defined icons, drawing with geometric shapes, and freehand drawing. Among a substantial number of QbD systems, we list the typical systems in each type of drawn query.

2.4.2.1 Drawing Query on Computer

The earlier studies on QbD mainly considered shape matching between the line drawings produced by users and edge features of database images. For example, the *Elastic Template Matching* (ETM) system [35] matched simple shape to a painting database (see Figure 2.46). By adding labels onto the outlined area, *MindFinder* combined QbD image search with keyword-based search [17], and users could also indicate the primary colour of a whole image in its improved version [17], see Figure 2.47. A common way to add colour information was to draw the colour line sketch directly, which was taken by Budd [16]. In

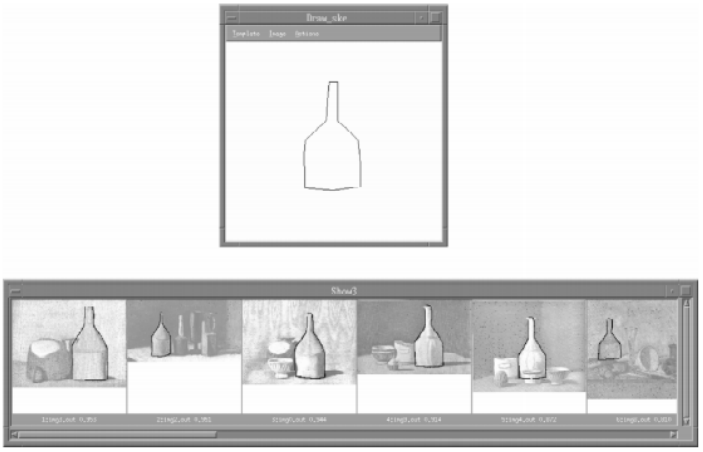


FIGURE 2.46: Screenshot of the user interface of ETM system (taken from [35]).

Budd's work, participants were asked to produce sketches by using of *Microsoft Paint*^{TM}. Three most typical sketching styles of users' sketches were recognised (see Figure 2.48): black-white line sketch, colour line sketch, and colour blob sketch. With the aim to evaluate the precision and efficiency of their novel algorithm, 50 sketches for each type were loaded into a *Query-by-Sketch* system and the top 20 closest-matched results were examined.



FIGURE 2.47: Queries of *MindFinder* system: sketch + tags (left), sketch + tag + colour (right) (taken from [17]).

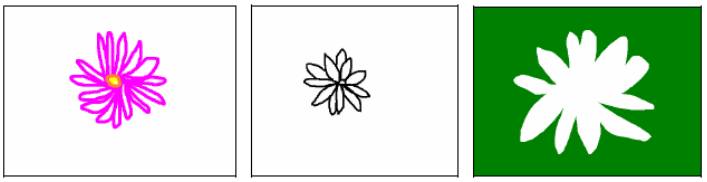


FIGURE 2.48: Three sketch styles: black-white line sketch, colour line sketch, and colour blob sketch (taken from [16]).

Icon-based drawing systems usually provide a limited set of iconic representations of general objects such as sky, trees, water etc., and users could create a

query intuitively by dragging or pasting icons to indicate rough spatial relations among these objects [84], see Figure 2.49 (left). Or users could directly draw lines to separate canvas into several areas and used semantic icons to represent the inner-region drawing, such as Hayashi et al.'s prototype [61] shown in Figure 2.49 (right).



FIGURE 2.49: The icon-based query formulation interface in Lew [84] (left), and a landscape with its icon-based query in Hayashi et al. [61] (right).

Pre-defined geometric shapes, such as line, rectangle, circle, filled shapes, etc., were used in the *VISI* system [65], as shown in Figure 2.50 (left).

Freehand drawing tools usually contain a pen or pencil and a colour palette with a limited number of colour choices (such as *Retrievr*) or a continuous colour picker, such as the fuzzy image search in *digiKim* and *ImgSeek* (please refer

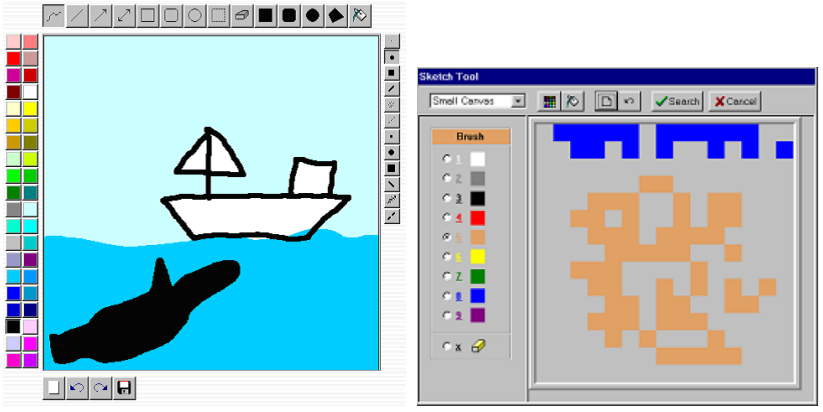


FIGURE 2.50: The pre-defined shapes of the *VISI* system (left, taken from [65]) and the sketch tool in the *CHROMA* system (right, taken from [80]).

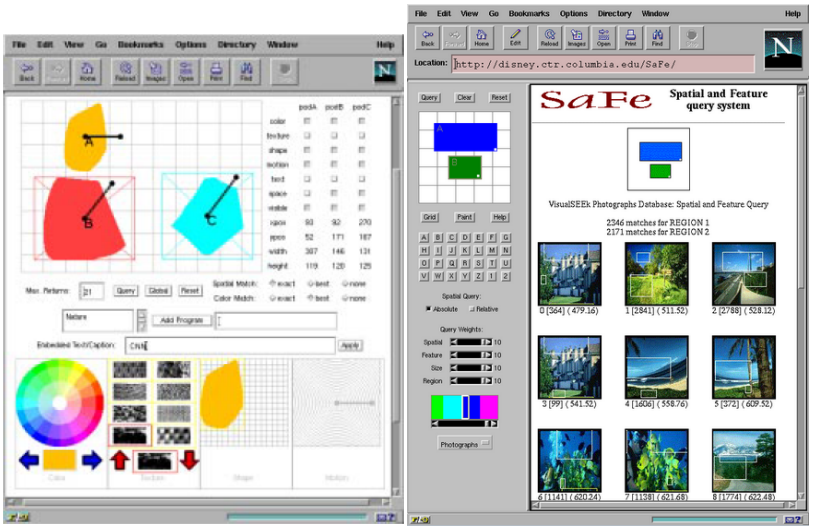


FIGURE 2.51: Screenshots of *VisualSEEK* (left) and *WebSEEK* (right).

to Section 2.2.3.1). One disadvantage of freehand drawing is that it is time-consuming [66]. Hence, some researchers examined the intermediate level of query formulation complexity between freehand drawing and icon drawing. The *CHROMA* system [80, 81] provides a grid-based drawing area for the user to fill colours in some squares to construct a colour layout as a query, as shown in Figure 2.50 (right). In both *VisualSEEK* system [122] and *WebSEEK* system [123], the user can construct a colour layout by sketching a region with different sizes and drag them onto the query grid. The user can also refine the properties of each region, such as colour, size, texture and relative location on the grid canvas, as illustrated in Figure 2.51.

However, the aforementioned systems are computer-based software or websites, and it is difficult to use a mouse to draw sketches on computer-based interfaces. As Korfhage [78] suggested, it would be more natural for the user to draw with a stylus on some form of graphics tablet instead of using a mouse. The increasingly use of touchscreens on mobile devices offers the user an intuitive way to use fingers to interact with drawing tools.

2.4.2.2 Drawing Query on Mobile Devices

Little previous work relevant to the *Query-by-Drawing* on mobile devices was found in the literature, and the majority of work aimed to explore the performance of image retrieval algorithms. Moreover, most systems use black-and-white line drawings as input query, such as the system in Tseng et al.'s work [131] shown in Figure 2.52 (left), rather than intuitive colour sketches. The *mCLOVER* [75] enables users to sketch contour of a leaf by dragging a stylus across the phone screen. Similarly, *SketchMatch* is a mobile game [125] built on *Windows* phones to match and rank user's simple line sketches based on both shape and position of strokes, as shown in Figure 2.52.



FIGURE 2.52: Screenshots of iPad user interface of Tseng et al.'s system [131] (left) and *SketchMatch* Windows App [125] (the rest).

2.4.3 Drawing Behaviour Study

A typical drawing process consists of a sequence of user actions. Blaser [11] asserted that some common drawing patterns existed because people could interpret other's drawings under certain context. The literature on drawing behaviour provides the fundamental knowledge about how people draw and what strategies they use.

The earlier work about drawing behaviour we could find is Bogles's Master dissertation [12]. Bogles reviewed previous drawing behaviour studies that mainly

focused on the individual's drawing development, such as from pre-schooler to early grade school children, over young adolescents and to adults. Bogles further investigated the characteristics of individual's drawing behaviour from five different age groups: preschoolers, six, nine, twelve-year olds, and adults. An empirical study asked 148 participants to draw a *flower* and a *dog* under three conditions: model-absent, model briefly-present, and model continuously-present. The results supported the literature that drawing behaviour changed with age and Bogles concluded that drawing performance was enhanced by continuously presenting the model. We are rather interested in the metrics that Bogles used to analyse the participants' drawings [12]: the *number of colours* used in each drawing, the *use of unrealistic colours*, the *percentage of utilised area* to make the drawing, the *background*, the *number of overlappings*, the *frontal/vertical perspective*, the *left-right* and *top-bottom sequencing*, the *time taken to complete a drawing*, and the *rated completeness score*.

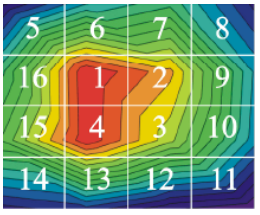


FIGURE 2.53: Spatial distribution of objects on canvas. Red indicates the highest and dark blue shows the lowest density of objects (taken from [11]).

The seminal work by Blaser [11] inspires us to explore users' drawing behaviour during query formulation on mobile phones. Blaser surveyed people's sketching habits by asking 32 participants with various backgrounds to produce one sketch for three tasks - a familiar scenario, an unfamiliar scenario and an imaginary scenario. Based on evaluation of 91 resulting sketches, Blaser concluded that people's sketches were simple and abstract, therefore the entire context was more important than single objects when interpreting a sketch [11]. Blaser also

recognised that people tend to draw incomplete objects under unfamiliar scenario, and most objects were drawn in the centre of a canvas, as shown in Figure 2.53, partition 1 and partition 4 indicate the highest object density.

2.4.4 Current Mobile Drawing Applications

At present, there are numerous mobile applications (Apps) that can be utilised for the purpose of drawing. As mentioned in Section 2.2.3.2, we have reviewed 140 Android Apps for drawing purpose. Most of these drawing mobile Apps share similar features, for example, the ability to:



FIGURE 2.54: Screenshots of *InfinitDesign* (left) and *CanvasLite* (right).

- modify the canvas backgrounds with plain colour or texture, e.g., 30% of reviewed Apps separated brush colour choosing from background colour choosing or texture-filling;
- change the size and colour of pen or pencil or brush (both limited or infinite number of size choices);
- select a colour from more than one colour pickers, e.g., 21.4% of reviewed Apps have two colour pickers, 12.1% of 140 Apps use three or more colour selection tools;

- choose from several types of pen or pencil or brush, etc., as shown in Figure 2.54.

Some drawing Apps specially offer the options to:

- use several stamps or icons to draw, e.g., *Sketch* shown in Figure 2.55;
- choose from '*Portrait*' and '*Landscape*' screen orientation before starting sketching, e.g., Figure 2.55 (left);
- draw on real colour photographs; and
- save and share drawings via social networks.

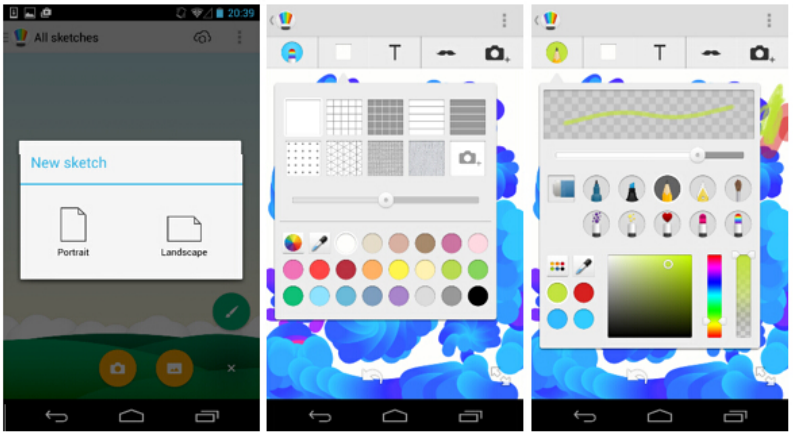


FIGURE 2.55: Screenshots of *Sketch*.

The input areas of continuous colour pickers of current drawing Apps are usually designed in the form of two-dimensional (2D) or 1D-2D. That is, one motion simultaneously changes two parameters of a colour and one motion modifies another parameter. Unfortunately, it seems that current drawing mobile applications usually give users a considerably large set of tools that the novice or non-professional users usually cannot use effectively [95]. Further, there is a paucity of applications relevant to drawing a query to search for images. We only found one App that explores line-drawing as a query in current App store

(there might be others not being reviewed), a Windows mobile phone game *SketchMatch* (please refer to Section 2.4.2.2). None of these reviewed mobile applications are specifically designed for creating a colour drawing to search for images.

2.5 Drawing from Memory

Note that this PhD work is not concerned with colour memory or memory drawing, however, it inevitably requires a short-term memory to store the visual form of stimulus when drawing on a surface without seeing it [94]. Moreover, our study of drawing colour sketches from memory will be benefited from previous research with respect to experimental design.

2.5.1 Colour Memory Study

The literature reveals that colour memory is complex. According to Seliger [116], there are three types of colour memory: short-term delayed matching and two types of long-term recall. In a colour memory experiment, participants are usually tested to match a stimulus after different delay periods [44, 59]. Besides colour memory itself, other factors might affect memory performance, such as the type of present stimulus, prior experience associated with observers [68], and the performed task itself. A frequently adopted task is to ask participants to adjust a colour until it matches the original colour sample [102, 116], sometimes participants have to pick a colour from a colour set [87] or to decide whether two colour samples are same or not [49].

To investigate how stable colour memory is, researchers studied the colour memory over different time delays. Rather short delays were adopted during the early

studies, e.g., Nilsson and Nelson [102] chose short delays from 0.1 to 24.3 seconds (sec). With a delay of either 0.1-sec, 1-sec or 10-sec, Francis and Irwin [49] asked observers to judge whether two samples presented with context are the same or not. Perez-Carpinell et al. [106] carried out a study with longer delays 15 seconds, 15 minutes, and 24 hours to select a Munsell colour chip that best matches the previously-seen colour, and found that the accuracy of colour memory decreased rapidly as the delayed time increased. However, Hamwi and Landis [59] found that there was no relationship between colour memory and delayed time, by using session delays of 15 minutes, 24 hours, and 64 hours. D'Ath et al. [29] investigated hue memory immediately after seeing the tested colour, after one hour and after one week.

Deterioration of colour memory has been found by many researchers. Memorised colours were more saturated and slightly brighter than the original ones [99]. Henry et al. [63] confirmed that people tend to select colours that are much more saturated than the original colours. Uchikawa [132] came to the same conclusion with respect to *saturation* by asking participants to judge the saturation of two colour stimuli. However, it is impossible to draw general conclusions about colour memory from the study with different design paradigms, e.g., colour stimulus, colour matching tasks or colour selection tasks, the number of participants, and different delayed periods. Hence, colour memory is still a controversial topic.

2.5.2 Memory Drawing Study

The task of drawing from memory is most frequently used to conduct diagnostic tests for Neuro-psychological impairments [129], such as constructional apraxia, aphasia, and visual agnosia, etc. Usually, drawing from memory tasks include the freehand drawing of familiar objects or simple copying task [85]. For example, Gainotti et al. [50] asked participants to reproduce line drawings



FIGURE 2.56: 12 brand logos studied by *Monochrom*.

from memory after showing the participants stimuli for a short time, and these stimuli have characteristic shapes (e.g., hand, table, arrow, or umbrella).

Some researchers have explored people's memory related to commercial logos and brand names. Ly and Hardeberg [88] examined long-term colour memory of well-known brands such as *Coca-Cola red*, *McDonalds yellow*, *Heineken green* and *Facebook blue*, with the aim to find out which colours are more memorable to customers. They asked 16 participants to reproduce the abovementioned four colours with context (filling in the missing colour on the logos) and without context (text clue only). *Monochrom*[23] attempted to evaluate the attraction and memorability of 12 commercial logos (shown in Figure 2.56) by asking Austrian people to draw logos from memory on a whiteboard. However, the analysis of drawings, e.g., the drawing accuracy, cannot be found on their website.

More recently, psychologists from UCLA [10] conducted an academic experiment to ask 85 participants to draw the *Apple logo* from memory on paper and also to differentiate the correct Apple logo from 11 distractors. They found that the frequent exposure of the logo may lead to the poor memory because people

[23] http://www.monochrom.at/markenzeichnen/index-eng.htm

do not have the intent or functional reasons to encode the details, which supports previous theories on *inattentional amnesia* [148] or *inattentional blindness* [118].

In summary, the stimuli used in the previous memory drawing studies are all objects-based, such as everyday-objects or logos in daily life, to which people are constantly exposed, so participants might pay less attention to the details of a stimuli due to availability. We are more interested in the situation that people have an awareness of the availability of a *Query-by-Drawing* application that allows the user to draw a query roughly to search for paintings, and therefore there is a functional reason that one needs to memorise colour information and the associated details in the painting.

2.6 Summary

In this chapter, we have presented a comprehensive and systematic review of the methods and techniques related to the investigation of QbD image search mobile Apps. We have reviewed the literature and mobile applications related to colour selection, colour picker, and Query-by-Drawing, and the relevant methodologies of studying colour memory and drawing from memory that could be used for our experimental design. We have also proposed a new way of categorise colour pickers and conducted a survey of drawing-based mobile applications.

Colour selection can be carried out with the facilitation of a colour picker built into the application. However, there is still no guideline about choosing a suitable visual representation of a colour picker for a QbD application. The current design of colour pickers is largely dependent on the developer of the application. As Ware [143] proposed, the problem of the best colour selection interface is by no means resolved.

To the best of our knowledge, there is little previous research related to Query-by-Colour-Drawing on mobile devices. According to Jrvelin [70], the performance of a CBIR system is affected by users, tasks and systems. Hence, to improve the efficiency of QbD painting search on mobile devices, it is important to thoroughly understand user behaviour and how factors relevant to the users, tasks, and interface design affect task performance and system usability.

Chapter 3

Introducing our *Query-by-Drawing* Mobile Application

3.1 Introduction

To the best of our knowledge, although there are many innovative content-based image retrieval (CBIR) techniques in the literature, little research focussed purely on improving the image search user experience through exploring and designing user-friendly user interfaces. The advent of touchscreen mobile devices provides an appropriate platform for Query-by-Drawing image search on mobile devices. Therefore, rather than investigating the image retrieval algorithm itself, we focus on exploring user-friendly mobile user interfaces that better suit the existing CBIR algorithms. Specifically, we adopt the *Fast Multi-resolution Image Querying (FMIQ)* method [69] as image matching algorithm because its colour features are effective for painting search and it is a fast technique for real-time search on mobile phone (Section 3.2.2).

In this chapter, we introduce our Query-by-Drawing application, including the system architecture (Section 3.2.1), the database (Section 3.2.3), the general user

interfaces (Section 3.3), and early colour picker interfaces (Section 3.3.1.1). A focus group study involving four human-factors experts is conducted to evaluate the usability of our preliminary user interfaces (Section 3.3.1) and the feasibility of line-drawing and colour-drawing query. Finally, based on the findings from the focus group, the refinements of drawing tool interfaces (Section 3.3.2) and colour picker interfaces (Section 3.3.3) are also discussed.

3.2 Overview of System

Our Query-by-Drawing application is designed to allow users to search for a painting or similar ones by roughly drawing colour sketches of their search ideas. The drawing tools are simple, and the user only needs to select one of five different sizes of brushes and pick a colour to draw rough colour sketches.

3.2.1 Client-Server Architecture

The application developed in this PhD project adopts the Client-Server architecture, as shown in Figure 3.1. The front-end of QbD system contains an online user interface module, which is implemented on Android mobile phones. The back-end consists of an online module that executes the matching of a user's drawing with the database and then sends a set of the most-matched results to the user's phone. Moreover, the back-end also includes an offline module that handles the storage of an image database and the pre-calculated features of database images. The server we used is Microsoft's Cloud service - *Windows Azure*[1]. To support real-time responses, the Internet connection between Client

[1] http://azure.microsoft.com/en-gb/

and Server is written in PHP[2] and the image retrieval algorithm is written in Python[3].

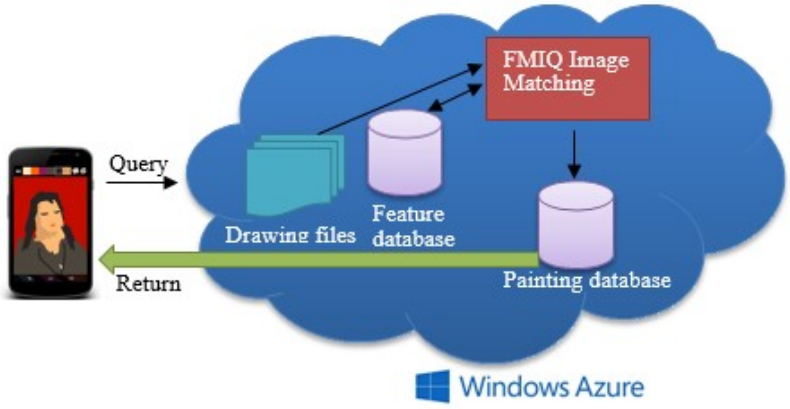


FIGURE 3.1: The Client-Server architecture of our QbD mobile application.

3.2.2 Image Matching Algorithm

According to Corridoni et al. [27], both colour and spatial arrangement of colour strokes are important for a painting. The *Fast Multi-resolution Image Querying algorithm* (FMIQ) proposed by Jacobs et al. [69] computes a colour feature description for the whole image. This method works pertinently fast based on image '*signature*' computed by *Haar wavelets decomposition*[4]. We briefly introduce the procedure of FMIQ as follows:

1. The database images are first resized to a resolution of 128 x 128 pixels and converted into YIQ colour space;

2. Only the largest Haar wavelet coefficients of each colour channel are saved as a signature of each image. Based on Jacobs' experimental results [69], the 60 largest coefficients in each channel work best for drawn queries and memory queries;

[2] http://www.w3schools.com/php/php_intro.asp
[3] https://www.python.org/
[4] http://en.wikipedia.org/wiki/Haar_wavelet

3. Quantise the coefficients to +1 (positive) or -1 (negative) and store them into a six feature arrays which are a combination of three colour channel (Y, I, Q) and signs (positive, negative), see Figure 3.3;

4. The querying step firstly computes the colour feature of the query image, the same steps as the preprocessing database images;

5. The matching step then calculates the distance between the features of a query image and those of each database images;

6. Sort the distance in an increasing order. The image with the smallest distance to a query is the most similar match.

Jacobs et al. [69] evaluated FMIQ algorithm on a database of 20,000 paintings and demonstrated that it is well suited to match coarse-scale versions of images to high detail versions of the same image. We implemented the FMIQ algorithm in Python and stored the Python codes on the server. Portrait painting queries are only matched against the portrait paintings database, while landscape painting queries retrieved from the landscape painting database. Both the top 60 coefficients and the matching weights (see Figure 3.2) suggested by Jacob's et al. [69] are maintained because they are obtained based on a scenario of using drawing as the query for painting search, which is the same as our application.

	Painted			*Scanned*		
b	$w^Y[b]$	$w^I[b]$	$w^Q[b]$	$w^Y[b]$	$w^I[b]$	$w^Q[b]$
0	4.04	15.14	22.62	5.00	19.21	34.37
1	0.78	0.92	0.40	0.83	1.26	0.36
2	0.46	0.53	0.63	1.01	0.44	0.45
3	0.42	0.26	0.25	0.52	0.53	0.14
4	0.41	0.14	0.15	0.47	0.28	0.18
5	0.32	0.07	0.38	0.30	0.14	0.27

FIGURE 3.2: The set of weights used in our FMIQ implementation (taken from [69]).

3.2.3 Database

For the building painting database, we only selected the paintings with well-recognisable colour blobs and a colourful background. A set of 250 portrait paintings downloaded from the *Tate Collection*[5] are used to test the matching algorithm at the early stage of this PhD work, and a database of 768 abstract paintings downloaded from the *Tate Collection* are also included and used for search during a practice session before the formal tasks. To extend our study, high-resolution scanned copies of 400 landscape paintings and 400 portrait paintings are downloaded from the *British Broadcasting Corporation (BBC) Your Paintings* website[6]. All paintings belong to the representational western paintings.

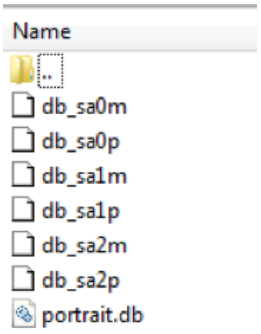


FIGURE 3.3: An example of portrait colour feature files stored in our Server.

All three databases, including the painting database, the colour feature database, and each painting's path files, are stored on the *Microsoft Azure* Server. For painting database, a separate folder is used to store each of three painting sets, including 250 Tate's portrait paintings, 400 BBC's portrait paintings, 400 BBC's landscape paintings, and 768 abstract paintings. Same as the colour feature database. For each painting set, colour features are store as file, as shown in Figure 3.3. The User's drawing files submitted to search are also saved on the server. The matched images will be downloaded from the database on our server

[5] http://www.tate.org.uk/about/our-work/collection/about-the-collection

[6] http://www.bbc.co.uk/arts/yourpaintings/

to the Client-phone after a user submitting a query. All database images can be accessed: http://min-horizon.cloudapp.net/

3.3 User Interfaces of QbD System

The interactions between users and the system mimic the physical drawing process of using a brush on a canvas, with the colour selection process enhanced by various colour pickers. The real-world painting process comprises two steps: adjusting brush properties and making strokes. The brush properties include colour, intensity, and wetness of the brush, among which adjusting the colour could be either by selecting an existing colour from a palette or by mixing more than two colours. A stroke is a step that made a footprint of the brush on the canvas.

We have developed our client applications based on an exploration of the user interface (UI) design of currently available drawing mobile Apps (refer to Section 2.4.4). However, our application aims to allow users to make a rough colour drawing rapidly, not for producing a precise drawing carefully. Therefore, only the basic drawing tools, including '*brush*', '*eraser*', '*colour palettes*', '*undo*', '*redo*', and '*delete all*', are provided. A focus group study is conducted to examine the general usability of our drawing application.

3.3.1 Focus Group Study on Early-stage of Prototype UI

The preliminary Query-by-Drawing application is implemented. As described in Section 3.2.2, Jacobs et al.'s FMIQ algorithm [69] is employed to match a colour drawing with the painting database.

The goal of a focus group study is to gain a basic understanding of the users' attitudes on drawing a query (colour drawing) of painting search as well as evaluating the usability of the user interface elements of our early prototype applications. The detailed report of the focus group study has also been published in a MM4CH workshop paper [153].

3.3.1.1 Early Prototype User Interfaces

The developed user interface allows users to draw on the front-end which interacts with our back-end system. The preliminary drawing interface includes brushes with five different sizes (5, 10, 15, 20 and 30 pixels), '*Colour picker*', erasers with five different sizes (5, 10, 15, 20 and 30 pixels), '*Undo*', '*Redo*', and '*Trash*' on the top of a canvas (see Figure 3.4).

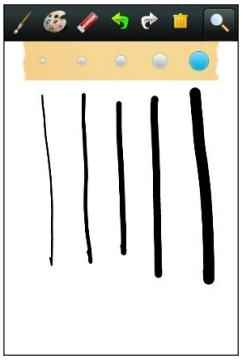


FIGURE 3.4: Screenshot of drawing tools of our early QbD prototype. Left to right: *brush, colour picker, eraser, undo, redo, trash*, and *search*.

The workflow of our Query-by-Drawing App is demonstrated in Figure 3.5. When the '*colour picker*' icon is tapped, a panel with six pre-defined colour swatches, a '*background colour picker*', and a '*brush colour picker*' will be shown on the top of the screen; when the '*background colour picker*' or '*brush colour picker*' icon is further tapped, the same colour ring will be shown on a dialogue, which consists of a *hue* ring and a *saturation-brightness* slider bar, and a small colour circular swatch in the centre of the *hue* ring. The colour ring

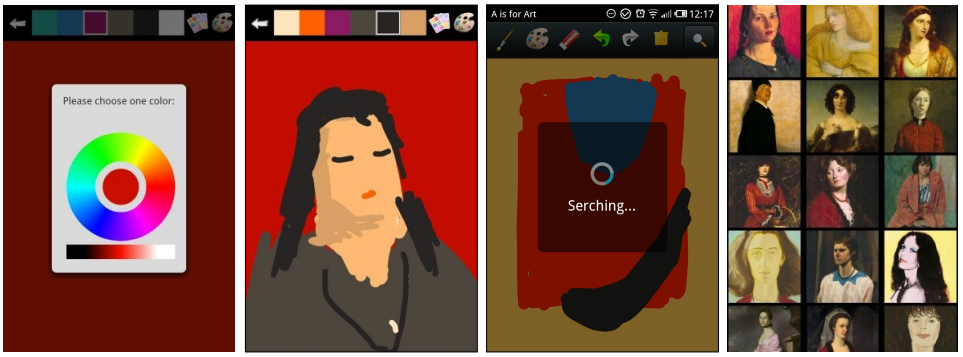


FIGURE 3.5: Screenshots of brush/background colour picker interface, work flow, and result-displaying user interfaces of preliminary Query-by-Colour-Drawing App.

picker is an extension of an APIdemos sample '*FingerPaint*' demo[7] that comes with the Android SDK.

When selecting a brush colour, the user can start with picking one of the six pre-defined colours, and adjust a colour by touching the *hue* ring or *saturation/brightness* slider bar. The central colour circle dynamically changes its colour according to the user's finger movements (e.g., the *red circle* shown in Figure 3.5, left). If the selected colour meets the user's intent, the user can press the colour circle in the middle of dialogue and the new colour will be automatically allocated to the brush colour and will be saved on the colour swatch that is selected. The user is also able to pick a colour directly on colour picker without a starting colour; in that case, the new colour will not be shown in the six-colour panel for later use.

When completing the drawings, the user can press the '*search*' button at the top-right corner. The 21 top-matched results will be displayed as thumbnail images with three in a row, as shown in Figure 3.5 (right).

[7] `Thisopen-sourceprojectwaslocatedin<sdk>\samples\android-<version>\ApiDemos\src\com\example\android\apis\graphics`

3.3.1.2 Participants

Four researchers from the University of Nottingham Human Factors Research Group voluntarily took part in the study. The levels of expertise are classified as the professionals in usability evaluation. All participants are novice users of our QbD App.

3.3.1.3 Procedure

The four experts (participants) were first introduced to the room and sit around a round table. Then the purpose of the focus group study was explained clearly to the participants and consent forms were signed before undertaking the formal tasks. Participants were asked to make themselves familiar with our QbD systems and then to complete two tasks:

1. *Draw six paintings printed on the task sheet and search for them*;
2. *Draw anything you would like to find by use of our Query-by-Drawing application*.

Participants were encouraged not to worry about their drawing performance (e.g., the time and quality of drawing). They were informed that they could ask the observer for help if any problems are encountered. No instructions on how to use our QbD application were provided. The participants could write down their comments on the interactions based on Nielsen's usability evaluation guidelines [101], to make sure they could remember the issues they came across for the group discussion afterwards. During the debriefing, the experts were asked to make any suggestions they might have for the improvement of our App. The focus group study lasted for about one hour.

The participants were recorded with a tripod-mounted video camera (Canon Legria HF S21) in the corner of study room and a web-camera (Logitech C210) hung on a laptop for the whole session, with the purpose of the later analysis.

3.3.1.4 Results and Discussion

92.8% of tasks were completed by drawing and searching for paintings. One participant failed to find the portrait painting shown in the second row in Figure 3.6.

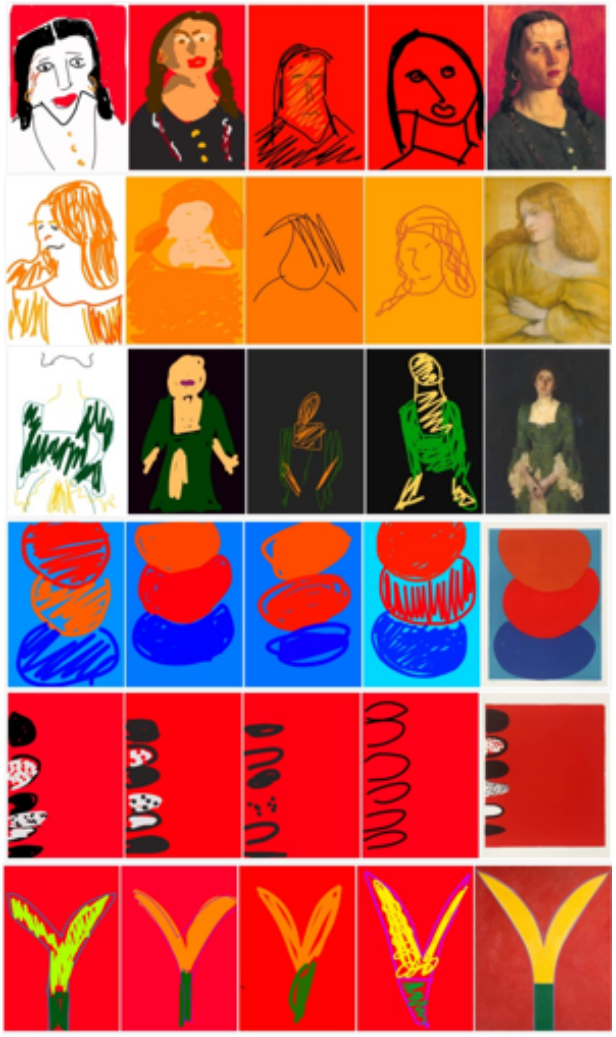


FIGURE 3.6: Colour drawings from four human-factor experts (drawings in one column from one person) and the original paintings (last column).

Although indeed there is a learning curve for the first-time user to explore the features of our QbD system, participants reported that they enjoyed using our system in terms of the idea of searching paintings by drawing. However, from the feedback gathered from both participants as well as our observations, there are some issues and design suggestions regarding the user interfaces:

- A colour selection procedure requires four steps, which makes it complicated for the users to pick a brush colour;
- The sizes of both brush and eraser are not big enough, and the eraser sizes should be bigger than the brush sizes;
- The icons of the brush colour picker and canvas colour picker are put closely and they are looked similar to each other, which might confuse the user;
- No tutorial about how to choose a colour is shown. The users do not know that tapping the circular swatch in the middle could select the expected colour;
- Participants would like to see more results than the top 21 most-matched images.

We also found the following behaviour from the participants' drawings and our own observations:

- The default settings of the drawing tools, such as the default brush size, are used more frequently by the participants.
- In colour drawings, the participants used hatching lines to fill colours. As shown in Figure 3.6, the hatching lines are drawn by using the brush with default size.

- The sketch of an object may not be placed at the right position and orientation at the correct scale. Also the perspective view of the sketched object might not be precise, e.g., users might draw the front view of a lady and the side view of a horse, which is in line with Vuurpijl et al. [139].

- The quality of the participants' drawings are improved by gradual refinements based on the search results. So there is a potential way to help the users to draw a query if the users knew how an image retrieval system works.

- Most of the experts prefer the thumbnail way of showing the retrieval results.

- The experts proposed that separation of brush colour picker and canvas colour picker avoids increasing the burden on the users.

We realise that the users usually cannot draw a precise shape of an object, such as the sketches shown in Figure 3.7, and one participant even gave up because she claimed that she could not draw the contour of aimed object. This finding is supported by Giangreco [52] who asked 11 participants to draw images by using provided tools, including tools to draw edges (line drawing), or colour sketch, or fuzzy-set operations, and found that all participants chose the colour sketches to represent their queries. Therefore, we decide to only use the colour drawing as query input for the further study.



FIGURE 3.7: The 'horse' drawings from experts in the focus group study.

We also found that the quality of videos recorded by the web-camera is quite poor. And same as the digital camera placed in the corner of the room, we can't re-observe the participants' interaction with our QbD system on mobile phone. Therefore, a screen-recorder is necessary to examine the users' drawing behaviour and drawing performance.

Based on the suggestions from the focus group study, the user interfaces are redesigned and rebuilt with the main refinements of the colour selection mechanism, the icons, and the sizes of brush and eraser. The revised user interfaces have been tested by three colleagues of the author, who are more satisfied with the revised application than the original version.

3.3.2 Revised User Interface of Basic Tools

The new user interfaces of general tools are implemented with self-explanatory icons indicating the functionality of the selected tool. For example, the icons of the brush colour picker and the trash bin have been changed, as shown in Figure 3.8. Both brush sizes and eraser sizes are increased, with each eraser size 10 pixels larger than the brush size respectively. The basic drawing tools include brushes of five different sizes (10, 20, 30, 40 and 60 pixels), '*Brush colour picker*', erasers of five bigger different sizes (20, 30, 40, 50 and 70 pixels), '*Canvas colour picker*', '*Undo*', '*Redo*', and '*Trash*' at the top of the canvas (see Figure 3.8, right).

We separate the brush colour selection and the canvas (background) colour selection to facilitate users to produce or to refine their drawings easily. The canvas will be fully filled with a selected background colour, and the eraser cannot remove the background colour but only the brush strokes. Secondly, we position the icons of brush colour picker and canvas colour picker on the tool panel

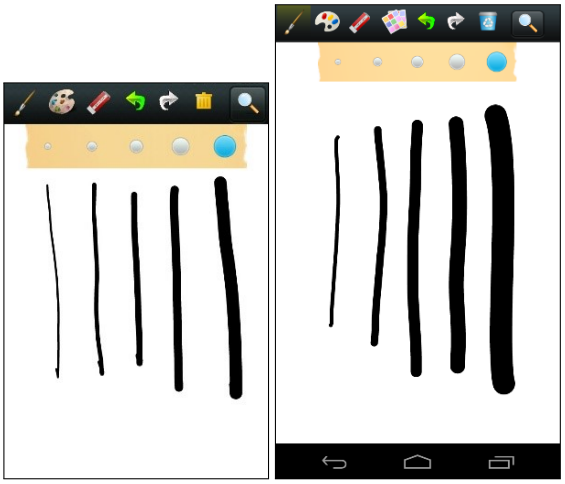


FIGURE 3.8: Screenshots of the initial (left) and the refined (right) drawing tools with self-explanatory icons. From left to right: *brush*, *brush colour picker*, *eraser*, *canvas colour picker*, *undo*, *redo*, *trash*, and *search*.

separately (see Figure 3.8, right) to avoid confusion and accidental pressing mistakes.

3.3.3 Revised Colour Picker Interface

It is impossible to implement and build all colour selection tools into an application. First, we decide to only focus on the HSV colour model (*hue-saturation-value*). There is a common belief that quoted from the standard reference work on computer graphics by Foley et al. [46]: "*the HSV model is user-oriented, being based on the intuitive appeal of the artist's tint, shade, and tone*." Dupplaw et al. [41] proposed that the HSV colour model is a compromise between the relevance to the human visual system and the computational effect. Another reason to adopt the HSV model in our work is because it is commonly used in the colour picker design of current mobile applications, as reviewed in Section 2.2.3.2.

Second, we choose to use the *square*, slider *bar*, *wheel*, and *ring* for the visual design (please refer to Section 2.2.2.1) of a continuous colour picker. Based

on Bogles's conclusions [12], all continuous colour pickers are designed in a combination of statically coloured mode and dynamically coloured mode. In other words, some elements are displayed fully statically, and some elements are partially dynamic, such as the *saturation* or *value* sliders or *saturation-value* 2-dimensional (2D) input area, because *saturation* and *value* must be represented with some hue. For example, the *brightness* slider shows a range of reddish colours when the user picks a red variant on the 2D input area, but the 2D input area will not adjust its colour when the slider is moved.

With different purposes of experiments, different colour selector interfaces are designed and implemented, e.g., the discrete colour swatches or the continuous colour pickers. We will introduce different colour picker interfaces with details in Section '*Colour picker interface*' of each experiment (see Chapter 5, Chapter 6, Chapter 7). We only describe here the modification of the preliminary colour picker interfaces based on the results of the focus group study.

The colour selection process is reduced to a two-step procedure: tapping on the icon of brush or background colour picker, and then picking the wanted colour. More feedback on user manipulation are shown to the user. For example, the '*OK*' is displayed in the centre of the colour circle that indicates the selected colour; and when the mixed colour is chosen, the colour ring with the selected colour will appear outside the colour circle, which informs the user that the new colour has been picked, see 3.9 (right). The context of the associated colour picker's icon will be changed into the new colour as well.

3.3.4 Result display Interface

A set of most matched images for a user's freehand-drawn query will be sent from the server to the user's mobile phone. The result display interface shows a group of small thumbnails of several most matched paintings for the user to skim

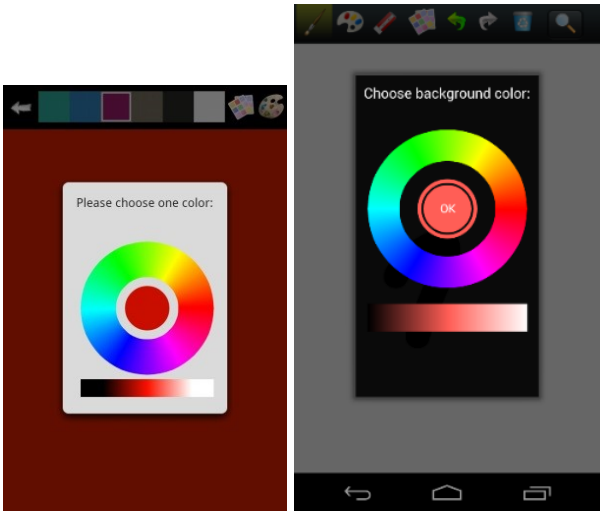


FIGURE 3.9: Screenshots of the initial (left) and the refined (right) colour picker interface.

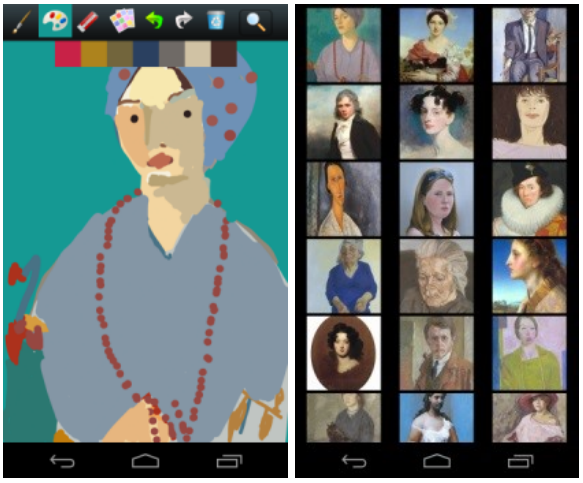


FIGURE 3.10: Screenshots of the final interfaces for a portrait query and result displaying.

over. Following the advice of Broek [15] as well as the mobile phone's property, a grid of 15 images per screen with three thumbnails in a row are displayed in descending order of the visual similarity to a query. That is, the most similar painting is at the top-left of screen (e.g., Figure 3.10, right). The user is able to scroll up and/or down to see other thumbnails and simultaneously a navigation bar will be shown on the right side of phone screen during sliding. The user can see the enlarged view of the thumbnail image by double-tapping on it.

3.4 Discussion

It is obvious that the image matching algorithm *Fast Multi-resolution Image Querying* is not 100% accurate, however, it is suitable for our Query-by-Drawing application. Because the FMIQ algorithm uses the primary colour information of an overall image. The image matching algorithm which considers the location of main colour features will be helpful for the cross-domain (drawing vs. scanned painting) Query-by-Drawing application.

3.5 Summary

This chapter has introduced an overview of the system design and implementation of a prototype of a drawing-based painting search system on mobile phones, including the system architecture, the implementation of principal techniques, and the user interfaces of drawing tools and colour picker.

Based on a focus group study, the capabilities and usability of two image matching technique are examined: Drawing-by-Line-Drawing and Drawing-by-Colour-Drawing. The Drawing-by-Colour-Drawing is determined to be used in further study. This prototype served as a proof of concept for Query-by-Drawing (QbD) image search, and it can also be used as a platform for exploring the effects of different attributes of colour pickers or other user interface elements on the usability of a QbD mobile application.

Chapter 4

Research Design and Methodology

4.1 Introduction

As proposed by Heimonen [62], the study of information (image) seeking should be specific to the user, task, and context. This concern is reflected by the methodologies and experimental design adopted in our research, which is to investigate and improve the user interface design of a Query-by-Drawing (QbD) image search mobile application and to examine the feasibility of a QbD image search paradigm used in real life.

In this chapter, we firstly introduce the overview of our experiments (Section 4.2), followed by discussing our general experimental design, which includes experimental planning (e.g., choosing stimuli) (Section 4.3), the methods used for conducting experiments (Section 4.4), and the methods for data collection and data analysis (Section 4.5). Some of these methods are used in all of our experiments, such as the participant recruitment, the experimental setting, interviews, questionnaires (Section 4.5.3.2), the methods for data collection and data analysis. Three experimental task types (Section 4.4.2) and the overall experimental procedure (Section 4.4.3) are also described. Finally, our research design

is justified (Section 4.6), followed by a discussion of some inevitable limitations (Section 4.7).

4.2 Study Overview

Within an experimentally controlled design, the experimental environment provides means for recording all user behaviour. We have conducted four experiments by asking participants to use multiple user interfaces to search for painting stimuli. These experiments can be expected to result in the guidelines for designing the user interfaces of a *Query-by-Drawing* (QbD) system on mobile devices. Both within-subjects and between-subjects design are adopted. Measurable outcome variables comprise the time of creating a query, the completeness of a query, the number of successful queries, the participants' usability choices, and human accuracy ratings (see Section 4.5 for details).

Each experiment is designed to gain a basic overview of answering one research question.

Experiment 1 aims to explore the effects of colour picker types and their combinations (including three typical discrete colour palettes and three continuous colour pickers) on the usability and performance of our QbD application. 37 participants took part in the one-by-one experimental sessions to draw the queries for portrait paintings. The preliminary analyses of the data collected from Experiment 1 have been reported in a MHCI conference paper [155].

In order to obtain an equal number of queries on each drawing interface and to examine the effects of task type, painting type and brush type on the usability and task performance, Experiment 2 is conducted. Similar approaches of conducting Experiment 1 are used except for additional tasks and brush type. The task type consists of drawing with target painting in sight, drawing from

the painting description, and drawing for shortly-present painting. Both portrait paintings and landscape paintings are included for each task type. Two new system interfaces have been designed and developed in Experiment 2 in order to investigate and compare the task performance by using pre-defined brush textures with using pure colours. The textures are extracted from database paintings with semantic meanings, e.g., sky, skin, etc. Eight different groups of texture palettes are selected from landscape and portrait paintings separately in our database, which are built into System 'Portrait' and System 'Landscape'. 38 participants performed three types of tasks by using three colour systems and two texture systems, and the effects of task type, painting type and brush type on the outcome measures are explored.

Based on the findings of Experiment 1 and Experiment 2, we conduct Experiment 3 to further examine whether the design attributes of 1 dimensional-2 dimensional (1D-2D) continuous colour pickers affects the usability and performance of our QbD App. The combinations and representations of three colour parameters based on the HSV colour model are explored, leading to three groups of comparisons: 1) combinations of three HSV parameters: *H-SV*, *S-HV*, and *V-HS*; 2) the shape of 1D input area: slider *bar* and slider *half-ring*; and 3) the shape of 2D input area: *square* and *wheel*. We have designed three *H-SV* colour pickers with a *bar-square* representation to compare colour parameter combinations; and the *V-HS* combination is used for two groups of representation comparisons, resulting in colour picker comparisons *Vbar-HSwheel* vs. *Vring-HSwheel* and *Vbar-HSsquare* vs. *Vbar-HSwheel*. 41 participants produced queries for three types of tasks by using five colour pickers. Preliminary analyses of this experiment can be found in an INTERACT conference paper [156].

Experiment 4 is a longitudinal memory drawing study with the aim to examine whether people could draw a query for previously-seen paintings after three

delayed periods - *15-minute* delay, *1-week* delay and *1-month* delay. 17 participants were involved in this longitudinal memory study, and seven participants are selected from participants of previous experiments. Statistical analyses are used to examine the differences in query formulation and query accuracy among three trials, as well as the effects of other potential factors on the task performance of different trials, such as painting type, drawing skill, and the experience with our *Query-by-Drawing* application. Preliminary results of this experiment have been published in an MOMM conference paper [154].

4.3 Experimental Planning

Prior to running each experiment, three individuals are recruited to take part in a mock experiment round to test the experimental design (e.g., system design, experimental setting, questionnaires, and task duration, etc.) and data collection methods. Due to confusion about some parts of the questionnaires, slight modifications of task description and questionnaires' wording are made. Digital and web camera are removed from the experimental setting, because of the uselessness and disturbance which is also found in our focus group study.

4.3.1 Determining Stimuli

To test a colour- and edge-based search algorithm, Giangreco [52] chose five photos as stimuli based on the complexity of the colours and edges in a photo. Similarly, we mainly consider the colours depicted in a painting when choosing stimuli, because the colour features of an image are used by the image matching technique built on our back-end server. All stimuli will be sampled from our paintings database. Before selecting specific painting(s), we have to decide which types of paintings are popular among the general public.

4.3.1.1 An Survey on Painting Preference

An online survey was conducted via *SurveyMonkey* to decide which type(s) of paintings should be chosen as stimuli. The overview of the survey has been introduced in Section 2.2.4, in this section we only focus on the analysis of 115 responses (80 males and 35 females) on the question "*which of the following painting types or themes do you enjoy the most when visiting a museum or art gallery?*" More than half of the participants (57%) indicate their artistic preference of realistic paintings (see Figure 4.1). The most liked subject depicted in a painting are landscapes (43.28%) followed by persons (22.61%), as illustrated by Figure 4.2. There is no significant difference in art preference between outdoor and indoor paintings, see Figure 4.3. With regard to landscape paintings, only 5.22% of responders prefer to see the cityscape painting, and rural landscape paintings are most popular (46.09%), while still 30.43% of the participants do not care about the subject matter depicted in landscape paintings. As for portrait paintings, most of the responses (66.96%) indicate no preference for the subject matter; the second most preferred (21.74%) are the portrait paintings of female characters.

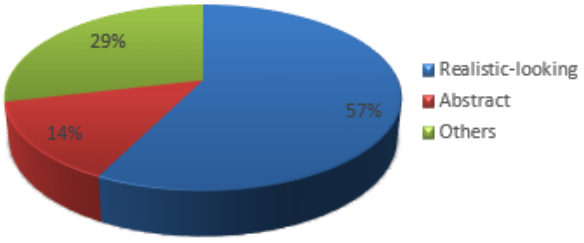


FIGURE 4.1: Distribution of the painting preference among realism, abstract or others.

Based on the survey findings, realistic landscape and portrait paintings are the first criterion for selecting stimuli.

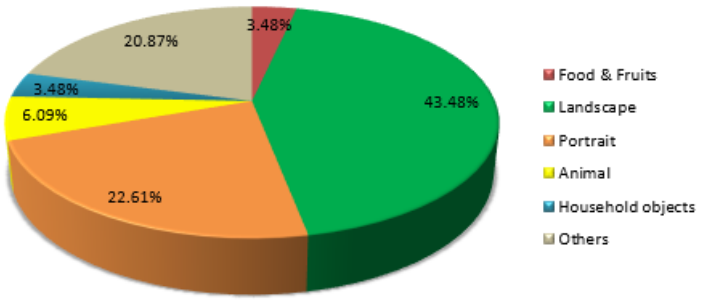


FIGURE 4.2: Distribution of the painting preference for different themes.

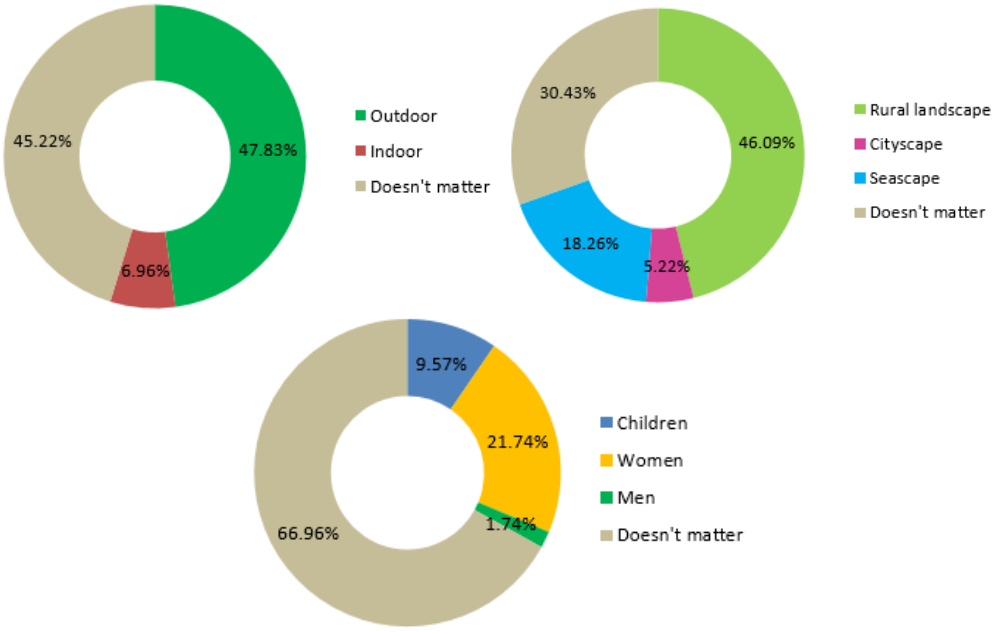


FIGURE 4.3: Distributions of the painting preference for different subject matter.

4.3.1.2 Stimulus Choosing Rationales

The choice of painting stimuli follows two rationales: first, the painting is less well-known; second, the painting should have at least four gist colours: *blue*, *red*, *green* and *yellow*, which are commonly used in the literature and described as the distinctive and basic colours [7] or focal colours [36]. It seems that researchers have agreed on that the number of colours should be limited to between four and seven distinct colours [124]. As an example, Figure 4.4 shows

an abstract painting used as the stimulus for practice session, which allows participants to understand how our system works but would not increase the participants' workload because it is simple to draw.



FIGURE 4.4: The stimulus used in practice session.

For target-unspecific task, we select everyday familiar objects or scenes such as green grass, blue sky, or skin tones, with which "*people have frequent visual experience and are likely to produce memory colours that are common to many people.*" [7]. With varied aims of each experiment, the specific stimuli are chosen correspondingly. The stimuli used in each experiment are detailed in the '*Experimental Design and Procedure*' Sections.

4.3.2 Participant Recruitment

The participants are recruited from the university and college through e-mail and advertisement posted on the whiteboards on campus. Participants are self-selected in terms of age, gender, background, drawing skills and colour experience. Most of the participants are European and Chinese. For the longitudinal memory study, the only prerequisite is that participants should be novice users of our QbD application. Each participant is compensated with five pounds Amazon voucher after completing each experiment.

Participants are not tested for *colour-vision deficiency* (CVD), but are asked whether they have CVD. We feel that the CVD screening will make the experiment more time-consuming and it has been warranted by Moretti [96] who

ask CVD participants to choose colours to create several colour schemes in his study. Those self-indicated CVD participants are not excluded, but their drawing processes and responses during the experiment are marked and analysed especially.

4.3.3 Ethical Approval

In accordance with the University policy, the experimental work conducted for this PhD research was approved by the University of Nottingham, School of Engineering ethics committee, with the approval No. 2013-39. The signed informed consent forms will be obtained from participants stating that they know all data is confidential and would be stored securely and would only be used for the purposes of research. A copy of the consent form can be found in Appendix A.

4.4 Methods for Experimental Execution

4.4.1 Experimental Setting

To avoid disruptions, all experiments are conducted in a quiet room with daylight, such as the University meeting room, offices, a computer terminal room. We decide to observe participant behaviour one-by-one in order to reduce the influence of peers, which means that only one participant conducts tasks during each experimental session. We assume that our experiment allows participants to perform tasks in a natural environment with one experimenter observing and note-taking.

Two Google Nexus 4 phones (133.9mm in height $\times$ 68.7mm in width) with Android 4.1 operating system are used: one phone is used to show the reference

paintings to the participants; and our QbD App and the screen recording App *SCR ScreenRecorder Pro*[1] are installed on the other phone on which participants would interact with our QbD App. Both mobile phones have been adjusted to the same settings such as the same level of brightness, to reduce any differences. We firstly attempt to use the digital camera on a tripod placed on the corner to record the whole procedure same as the focus group. However, we found that the participants acted unnaturally in front of the camera. Alternatively, the relative small *Ziggi*[2] *USB* camera, shown in Figure 4.5, is adopted to capture the participant interactions with the screen of mobile phone. It is reported by the participants that the Ziggi USB camera enables them to immersively engage in the interactions with our mobile application. The camera recordings are synchronised with the screen recorder made through all experimental sessions.



FIGURE 4.5: The setting of our experiments and the *Ziggi USB* camera (right).

4.4.2 Experimental Task Type

Three types of tasks are recognised when considering real-life scenarios of our Query-by-Drawing application, including:

[1] https://play.google.com/store/apps/details?id=com.iwobanas.screenrecorder.pro&hl=en_GB

[2] http://www.ipevo.com/prods/IPEVO-Ziggi-HD-USB-Document-Camera

T1 task Drawing with *target continuously-present* (also referred to as *copying-task*) - the reference painting remains shown during the whole drawing process and result browsing. Participants draw on one phone with the other phone showing the target painting.

T2 task Drawing with *target absent* (also refer to as *drawing from the description*) - T2 task is the target-unspecific task. For portrait paintings, participants are asked to draw from their imagination based on the English description of target painting. For landscape painting, participants are required to form a query of four seasonal landscape paintings, with each one is inclusive of at least five objects: sky, mountain, river, tree, and land. At the end of each query session, participants have to tell the examiner where the similar painting is located in the returned search results.

T3 task Drawing with *target shortly present* (also referred to as *drawing from 30-second viewing task* - one reference painting is removed from sight after 30 seconds, then participants are asked to draw this painting immediately after viewing. Participants have to judge if the target painting is in the returned results or not. The viewing time up to 30 seconds is also adopted in Laws [82].

For target-specific tasks (T1 tasks and T3 tasks), it is worthwhile to note that the reference portrait painting is shown in the portrait screen orientation while the landscape painting is presented in the landscape screen orientation. The reference phone and the drawing phone are adjusted to the same settings, such as the same level of brightness. The task of drawing from the description (T2 task) stimulates the common search in reality, because people usually look for the title or artist of a previously-seen painting by describing objects and colours in the expected painting.

People 'see' a colour differently, especially for drawing a colour query for target-absent tasks (T2 tasks and T3 tasks) which require people's mental imagination [111]. The errors raised from participant perception and cognition of a colour description or colour memory may severely affect the participants' drawings and consequently the interface comparisons. While in the copying-task (T1 task), participants are able to continuously look at the reference painting during the drawing process, which makes the different interfaces of drawing tools comparable somehow. Therefore, we ask each participant to conduct one copying-task on all systems with different designs in a random order, and participants' task performance and preference of colour picker interfaces are evaluated and compared.

T2 tasks and T3 tasks aim to provide a set of real-world search tasks that fully engage participants with a certain degree of interactions with our system, and thus we could examine the participants' ability to draw a query by using our QbD application and the participants' attitudes towards the search for paintings by drawing colour sketches. Participants produce a query for a target-unspecific task by using one system only, and participants will be asked to speak out their mind states, e.g., what they are trying to do, or some troubles they come across (please refer to Think-Aloud protocol described in Section 4.5.3.1). Participants are informed that they could produce as many queries as they think necessary.

4.4.3 Experimental Procedure

All experimental procedures consist of several one-by-one or two-by-one sessions involving only one or two participants and the thesis author who observes and takes notes. Each individual experimental session follows the same steps. We guide the participants through the aims and tasks of experiments, and then the participants will be asked to give their written informed consent for the

experiments. Participants are instructed to complete the pre-questionnaire followed by a drawing practice session when participants are allowed to familiarise themselves with the functionalities of our QbD App. After participants feel confident enough to use our App, we ask them to complete the actual tasks. Participants are told to perform the tasks in a manner as natural as possible. The post-questionnaire is presented during the whole experimental procedure, in which the participants are asked to rate on several topics related to system usability and to take notes as clues in an exit interview, if any.

After completing all tasks, a short exit interview will be conducted to allow the participants to express their general opinions on both user interfaces and their engagement with our application. All procedures will be screen-, audio- and video-recorded.

4.5 Methods for Data Collection and Data Analysis

The measures could be divided into objective measures, such as the time of drawing a query, and subjective measures, such as human accuracy ratings. The main performance indicators are generated from the post-questionnaires (please refer to Appendix A), video logs and the actual search results recorded by the examiner. The assessment of each drawing is done manually and intermediate results are stored in *Microsoft Excel* spreadsheets.

4.5.1 Measurements Definition

The outcome measures are collected from the drawing process of all experiments: the time spent in drawing a query, the *completeness* of query (including the *number of colours* and the *number of objects* depicted in a query), the *drawing accuracy* (including the 'objective' search performance the number

of successful queries, and the 'subjective' accuracy ratings), and participants' choices of *usability and preference*. These measures are chosen as the outcome variables because they are predominant features discussed frequently in the literature. The task performance measures include the time taken to complete a query, the number of successful queries, the completeness and the accuracy of a query. The number of colours drawn in a query is an indicator that the number of colours that users have selected, together with the number objects depicted in a query, these two measures reveal the completeness of a query and the ease of use of a colour picker. Given the knowledge of our colour-based QbD App, if a user uses fewer colours to draw more objects, for example, we might conclude that the colour picker does not facilitate the user to draw a query.

- **The time of query formulation** is defined as the amount of time since the participant starting to interact with our QbD application until the last move of creating the drawing.
- **The number of colours depicted in a query** is the number of unique colours in a drawing (counted by the experimenter).
- **The number of objects depicted in a query** is the number of unique objects with semantic meaning displayed in a drawing (counted by the experimenter).
- **Search performance** is obtained from the system's match algorithm. In our experiments both the *number of Top21 queries* and the *number of Top51 queries* are used to represent the search performance.
  - **The Top21 query** is defined as the query that hits the target painting in 21 top-matched images;
  - **The Top51 query** is defined as the query that hits the target painting in 51 top-matched images;

- Subjective accuracy measures are collected from human evaluators, including:
  - **Foreground colour accuracy** is the score ranked by human evaluators on the area of drawing created by using the brush together with a brush colour picker.
  - **Background accuracy** rating is the score assigned by human evaluators on the colours selected by using a background colour picker.
  - **Query accuracy** rating is the score rated by human evaluators on the entire query compared to the target painting.
- **Usability and preference measures** of colour picker interfaces are coded from the post-questionnaire into the ordinal data, which reveal participants' relative opinions on the individual colour selection interface, drawing process and searching experience.

4.5.2 Independent variables

Independent variables include the system design, colour picker type, task type, painting type, brush type, the visual design of colour picker, the delayed-time point and participant characteristics such as gender, ethnic group, drawing skills, and the experience with Query-by-Drawing. Different experiments emphasis on the investigation of the effects of different factors (variables) on the outcome measures depending on research questions.

4.5.3 Data Collection Methods

4.5.3.1 Think-Aloud Protocol

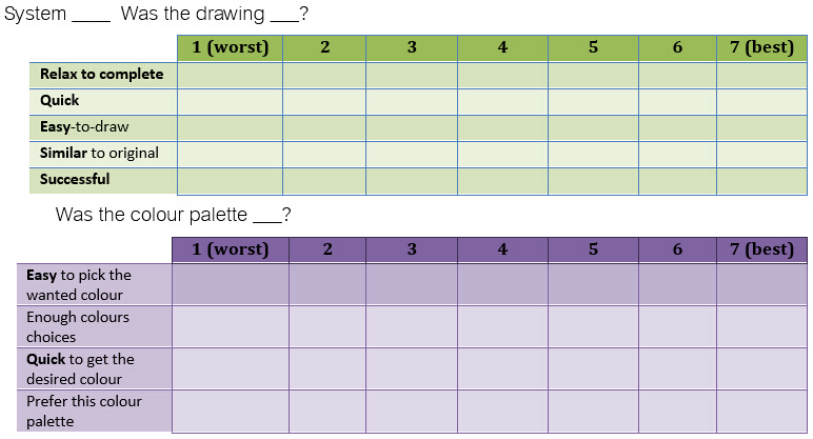
We ask all participants to verbalise what they are trying to do when performing all target-absent tasks (T2 tasks and T3 tasks). All of the participants' output and comments are recorded for further analysis. The use of the Think-Aloud method encourages the participants to express their confusion and the problem-solving thinking in incidental situations. Viable conclusions about usability and effectiveness of our QbD system are drawn from both participants' performance and verbal responses.

4.5.3.2 Questionnaires

Both pre-questionnaire and post-questionnaires are designed to allow for direct capture of participant's self-rating of their drawing skills, task performance and the usability of user interfaces. The pre-questionnaire is used in all experiments included in this thesis. The post-questionnaire with different topics and open questions are allocated to participants during different experimental sessions.

Pre-Questionnaire Demographic data such as the participants' age, gender, cultural background, the experience of QbD image searching and drawing skills, are collected from the pre-questionnaire (named **Q1**, see Appendix A). This information could be used to classify different participant characteristics.

Post-Questionnaire As suggested by Albert and Tullis [2] that task success might change the user's experience with a task. Therefore, all usability-related


System ____ Was the drawing ___?

	1 (worst)	2	3	4	5	6	7 (best)
Relax to complete							
Quick							
Easy-to-draw							
Similar to original							
Successful							

Was the colour palette ___?

	1 (worst)	2	3	4	5	6	7 (best)
Easy to pick the wanted colour							
Enough colours choices							
Quick to get the desired colour							
Prefer this colour palette							

FIGURE 4.6: One unit of the Questionnaire Q2.

ratings in the post-questionnaires should be conducted before participants viewing search results. The post-questionnaires mainly comprise several 7-point Likert scales to collect participants' opinions on their interactions with our QbD system in a quantifiable way, with a compensation of open-ended questions that allow the participants to specify their reasons of colour picker preference and the encountered problems. The granularity of 7 is chosen to offer participants an adequate space to differentiate ratings and also reduce participants' workload on the judgement. The post-questionnaires include several parts for different types of tasks introduced as follows:

Q2 : Measurements of system usability and task performance will be ranked by participants from 1 (extremely negative) to 7 (extremely positive) after completing a T1 task using one system, see Figure 4.6.

Q3 : Based on comparisons of different colour picker designs, participants are asked to produce the *best-worst scaling* [108] in terms of the effort, time, ease-of-use, enough colour choices, and preference after completing all queries for one stimulus under T1 task conditions, as shown in Figure 4.7. Screenshots of the brush colour picker and the background colour picker of each system are presented below the table on **Q3** as a reminder in case that participants forget the previously-used system interfaces (see

Compare system ABC and choose the one that:

	System A	System B	System C	
you spent **least effort** on drawing the query (✔)				Spent **Most** effort on drawing the query (✘)
you spent **less time** to get the wanted result (✔)				you spent **most time** to get the wanted result (✘)
Easiest to pick wanted brush colour (✔)				**Most difficult** to pick wanted brush colour (✘)
Easiest to pick wanted background colour(✔)				**Most difficult** to pick wanted background colour(✘)
Has **enough** choice of brush colours (✔)				Has **least** choice of brush colours (✘)
Has **enough** choice of background colours (✔)				Has **least** choice of background colours (✘)
you **like most** to pick brush colour (✔)				you **dislike most** to pick brush colour (✘)
you **like most** to pick background colour (✔)				you **dislike most** to pick background colour (✘)

FIGURE 4.7: Questionnaire Q3.

Appendix A). Previous researchers have demonstrated that the best-worst scaling is more efficient than traditional 'pick-one' choice [108].

Q4 : After completing each T2 task or T3 task, participants are asked to list the reasons for their system choice and also to indicate the level of agreement (1=strongly disagree, 5=strongly agree) on several statements (see Figure 4.8).

1) How much do you agree or disagree with the following statements? Please tick below

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I had a mental image of a painting that would satisfy my requirements.					
My query was an accurate representation of the type of paintings I had in mind.					
I am very happy with the painting(s) I found.					
I believe that I have seen all the possible paintings that satisfy my requirement.					
I believe that I have succeeded in my performance of the design task.					

FIGURE 4.8: One unit of the Questionnaire Q4.

Q5 : At the end of the experiment, participants are requested to rank their overall attitudes on a scale from 1 (extremely negative) to 7 (extremely positive) on the user experience of our QbD App, e.g., reliable, useful, novel, and fast etc., see Figure 4.9.

2) Was the Mobile App ___ ?

	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	
Satisfying								Frustrating
Reliable								Unreliable
Simple to use								Complex to use
Novel								Standard
Stimulating								Dull
Fast								Slow
Useful								Useless

FIGURE 4.9: Questionnaire Q5.

Open questions : Participants are asked to list the advantages and disadvantages of each system, as well as their suggestions for system improvement.

4.5.3.3 Interview

According to Ellis and Dix [43], the most relevant insight arises from user comments rather than various kinds of quantitative measures. The exit interview aims to get users' retrospective self-report feedback on the task performance, the system use and search experience, and some issues that are not covered in the questionnaires, such as the reasons for selecting the 'best colour picker'. The interview allows us to gain insights into what key factors are generally thought by participants on colour selection and drawing process.

4.5.3.4 Search Video Logs

As mentioned earlier, all experimental sessions of participants' task performing procedure are captured by *Screen Recorder Pro*, an application that could record participants' on-screen activity, e.g., the drawing process, the task completion time, the search results and the result images visited. The recorded videos and

audios are transcribed and provide rich user interaction data for qualitative data analysis of the drawing pattern and search strategy. The *Think-Aloud* data is also audio-recorded and transcribed into text files with the major problems numbered, which are used to understand participants' behaviour. Some transcripts are in Chinese, some excerpts quoted in the thesis are translated into English by the thesis author.

4.5.3.5 Drawings Ratings

The drawings submitted for results searching will be saved as a JPEG file on both phone and our Server. These drawings are uniquely named by the timestamp when finishing a drawing. Hence, we have to sort out drawings from each participant in a unique folder named with participant ID after each exit-interview. The advantage of drawings analysis is that drawings directly indicate participants' attempts to formulate their search intent.

Human Accuracy Rating As suggested by previous studies such as [18], the drawing accuracy is also assessed by human evaluators in terms of foreground colour, background colour, and overall colour impression of a query. In order to ensure the quality of the human evaluation work, the total drawings' scores are coded independently by three trained evaluators who are blind to the research motivations of this PhD work. The evaluators are also blind to the identity of participants who produce each drawing. The drawings for all tasks are presented to evaluators in a randomised order. The accuracy ranking data of drawings from each experiment are analysed separately.

Blaser's conclusions [11] that it is impossible to perceive the meaning of a drawing or sketch taken out of its context is most important for evaluating the 'drawing from description' task (T2 task) because no reference painting is provided to the evaluators. Therefore, we asked all evaluators to read our task descriptions

and they agreed on the interpretation rating criterion. Together with the participants' drawings, the original paintings and the written criteria on how accuracy should be rated are sent to each evaluator. In comparison to the ground truth painting, evaluators quantified the accuracy of each drawing by scoring three parts: background colour, foreground colour, and overall colour accuracy or overall query accuracy in the memory study (Experiment 4). Ratings are made on a 5-point Likert scale, with 1 representing very low accuracy and 5 representing very high accuracy. Since we have not found a difference among the three evaluators' ratings (reliability test, Cronbach's alpha=0.98) for the same group of drawings, we simply average the evaluators' ratings to provide the final accuracy score of each drawing in all tasks. Next, we will list the agreed criterion on how to differentiate the foreground and background of a drawing.

Coding Portrait Drawing The definitions of 'background' and 'foreground' are slightly different from the traditional meaning of a painting, which dedicate to describe the depth perception and perspective. During the coding of portrait drawings, the colour chosen via a background colour picker or the textural brush strokes that are used to fill out the whole canvas, are identified as the 'background'. The remaining parts of the drawing belong to the 'foreground', as illustrated in Figure 4.10.

Coding Landscape Drawing Similar to portrait drawing, the background of a landscape drawing is defined as the colour selected by using a background colour picker, and the foreground is created by using brush (e.g., Figure 4.11, left). However, the differentiation of the background in a landscape drawing is more of the challenges. First, the 'foreground object' might be represented by the background colour, such as the drawing shown in Figure 4.11 (middle). Second, the extreme case is that the whole drawing is made of foreground objects and no background is depicted at all, see Figure 4.11 (right). In contrast, the



FIGURE 4.10: Differentiation of background and foreground in a portrait drawing.

foreground is much easier to recognise, which is the area composed of all brush strokes (see Figure 4.11 for example).



FIGURE 4.11: The examples of coding the 'background' and 'foreground'.

4.5.4 Data Analysis Methods

Both quantitative and qualitative methods are applied to analyse the collected data in this thesis. All numeric data extracted from the questionnaires, video logs and human-ratings of accuracy are analysed by using the statistical analysis software *Statistical Package for Social Science* (IBM SPSS version 22.0). Unless

otherwise stated or defined, we use a significance level of α=0.05. The *Generalized Estimate Equation* (GEE) model [86] and *Generalised Linear Mixed Model* (GLMM) [126] are adopted to analyse relationship between the factors and outcome measures, and the '*backward model selection*' approach is adopted for the model fitting process by starting from the full-factor model and independent variable with highest non-significant p-value is excluded until the model having the lowest *Akaike Information Criterion* (AIC; [31]).

The GEE model is adopted because: 1) unequal numbers of participants between the stimulus group; 2) different numbers of query observations made on each of four systems (e.g., non-response data); 3) multiple measures from same individuals may cause correlation among the data; and 4) unequal variances in our samples (violation of Levene's test for homogeneity of variance, $p<.05$). In order to

The details of data analysis will be discussed in the 'Results' section of each experiment. The qualitative measures, which are collected from the interviews, the open-ended questions, the Think-Aloud data and the observation notes, are examined to produce valuable insights on interpreting the use strategies of performing different types of tasks on our Query-by-Drawing (QbD) application and also on inducing the current challenges for users and potential improvements of our QbD application.

4.6 Justifications of Research Design

As Kok [76] pointed at, there are many variables relevant to a valid experiment, such as the chosen stimuli, the experiment conditions, participants, and interface types, etc. In our experiments, we feel that the selections of both colour picker interface and the painting stimuli are all well-balanced, which are made based on the reviews of current drawing mobile applications and the survey results of the

general public's painting preference. The validity of the overall task design is also demonstrated by the study design in previous research [16, 32, 47, 53, 66].

Our QbD application is designed for general users from a wide population, and the goal of this PhD research is to draw general conclusions from experimental data about the behaviour of real users in realistic use scenarios [40]. The participants in our studies consist of students and adults recruited regardless of academic status and demographic background (e.g., gender, age, ethnicity and drawing skills), who are inevitably included in the target user population, and thus it is justified that our participants represent our target users [114].

The thesis author self-monitored all experimental sessions with one participant or two participants performing tasks at the same time, which reduces the possibility of biases caused by multiple experimenters and participant interaction.

4.7 Limitations

Some uncontrollable variations, such as light, mobile device bias, the time of day and the status of a participant when performing the tasks, might inevitably affect the experimental performance. Take the device biases for example: different screens of mobile devices (even with the same brand and same model) have different colour gamuts, which might influence the participants' colour perception when selecting a colour. Second, handheld web environments, e.g., the wireless broadband of smartphones, are still severely limited, which may affect the user experience of our QbD App.

With respect to participants, they might be interrupted by the monitor during task performing, e.g., answering observer's questions. Their opinions might also be affected by the observer's way of asking questions during the interview

and the knowledge of what the experimenter expects, known as *Response Bias*[3] or *experimenter demand effects* [110]. We cannot avoid the influence of these factors, but we can reduce their impacts, e.g., the monitor should keep quiet during the participants' drawing procedure.

Cognitive biases might happen among human evaluators who rate the drawing on the basis of likeness to the source painting [107] and their own interpretation and background, although they are trained before the rating work. That is the reason for having three evaluators to make judgements on each drawing.

Due to the time, participants, and resource, it is impossible to implement and evaluate all Query-by-Drawing image matching algorithms. Hence, we have to fix our attention on only one algorithm, the *Fast Multi-resolution Image Querying* (FMIQ) [69], which is obviously not 100% accurate, but it allows a real-time response for image search on mobile devices with limited computational capacity.

4.8 Summary

The overview of the experiments is firstly introduced in this chapter. The experiment planning, the methods used for experiment conducting and for data collection and data analysis are also described. The results of a survey about people's painting preference are used as the rationales for choosing paintings stimuli. The specific experimental task types and experiment procedures adopted through all experiments are briefly introduced, followed by defining the possible-affecting factors and outcome measures. We continue by discussing the justification of the methods employed in this thesis work, and this chapter ends with some inevitable limitations of our experimental design.

[3] http://en.wikipedia.org/wiki/Response_bias

We have to declare that the goal of this PhD research is to collect empirical data about how the factors affect the task performance and usability of Query-by-Drawing App when asking participants to perform specific tasks. Instead of bringing the participants into a lab, all our experiments will be conducted in a meeting room or office, which is a casual and real-life environment for potential users.

Chapter 5

Experiment 1 - Discrete and Continuous Colour Picker

5.1 Introduction

We have briefly discussed a series of colour pickers in Chapter 2, however, these are only examples of more than a thousand alternative designs on the current application market. Due to the limitations of time, participants and resources, it is clearly not possible to study all colour selector interfaces. There are two types of colour pickers: *discrete* and *continuous* (please refer to Section 2.2.2.3). We are interested in exploring the effects of these two colour picker types and their combinations on the usability and task performance of our *Query-by-Drawing* (QbD) application. Experiment 1 was carried out with 37 participants drawing queries for different tasks. Two typical discrete colour pickers and two representative continuous colour pickers are carefully designed and built into three systems: *discrete-discrete*, *continuous-continuous*, and *discrete-continuous*. Subjective measures show that the participants preferred the means of searching a painting by drawing rough colour sketches. The continuous colour picker was perceived

as more efficient and satisfiable by most participants. This experiment and the preliminary analyses have been published in [155].

In this chapter, the purposes of Experiment 1 are introduced in Section 5.2, followed by describing the colour picker interfaces (Section 5.3) and the experimental design and procedure (Section 5.4). In Section 5.5, we first present the statistical analysis of several factors' effects on task performance and participants' usability and preference choices when performing copying-tasks (Section 5.5.1). We also introduce the participants' query drawing patterns (Section 5.5.2) and query refinement behaviour (Section 5.5.3) and the search strategies (Section 5.5.4) used during task performing. Participants' attitudes on the overall system and suggestions of App improvement are described in Section 5.5.5 and a general discussion presented in Section 5.6.

5.2 Aims of Experiment

The primary aim of Experiment 1 is to compare the various measures (e.g., the time, accuracy, and preference) of drawing a query by using different colour picker interfaces. We also aim to get an insight into how people use our QbD application and what their opinions on the overall system are. The specific objectives are:

- To determine the colour picker and system design that people are most comfortable with, and to explore usability differences between continuous and discrete colour pickers and their combinations;
- To explore if personal characteristics such as gender, ethnic group and drawing skill affect the use of colour pickers and usability of the QbD App;

- To collect users' opinions about using our system including the user interface, the results returned by the system, and also suggestions for improving our QbD App;
- To see and understand how users draw queries and search on a mobile phone, and to use this information to redesign the QbD App's user interface.

Experiment 1 aims to answer the following research questions (RQ):

RQ1 : How do continuous colour picker, discrete colour picker and their combinations affect user's task performance, preference, and usability of mobile QbD painting search application?

RQ2 : How do users use different user interfaces to draw queries for different types of tasks? What do users feel about painting search by drawing colour sketches on mobile phones?

5.3 Colour Picker Interfaces

5.3.1 Database and Results Displa

The early-stage experiment in this PhD project focusses on portrait paintings only. The 250 portrait paintings selected from *Tate Collection*[1] are used as database images. The first 21 top matched images are returned to the user's phone from our Server.

[1] http://www.tate.org.uk/about/our-work/collection

5.3.2 Discrete Colour Pickers

The design of the discrete colour palette depends not only on the appropriate size of a colour set but also the colour values and their presentation. When examining the discrete colour palette, it is necessary to limit the number of colours, the colour sets, and the display style. To keep this PhD work in scope and also to make the duration of experimental session feasible, 256 discrete colours are adopted in this experiment.

As mentioned in Section 2.2.3.2, some existing colour palettes are extracted from the artworks of famous artists, e.g., the *van Gogh* colour palette in *Magic-Doodle*. Similarly, in our work we firstly use a *k-means clustering*[2] algorithm to extract the 256 dominant colours of 250 portrait paintings in our database. The k-means clustering approach clusters the pixels of database images in groups based on their HSV colours, and the centre of those clusters are the dominant colours of database images. The resulting 256 colours in a sequence of *k-mean clustering* output (see the hexdecimal value of each colour in Figure B.1 in Appendix B) are used directly as System A's background colour palette, with seven colour swatches in a row, as illustrated in Figure 5.1.

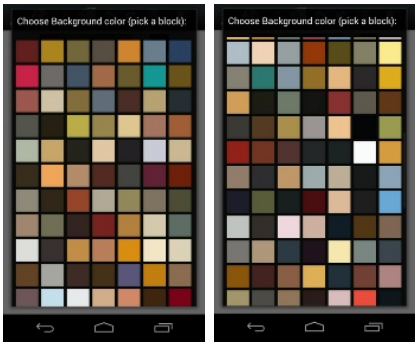


FIGURE 5.1: Screenshots of System A's background colour palette for portrait painting search in Experiment 1.

Next, we asked two human observers to categorise the 256 colours into seven colour groups: *red*, *yellow*, *green*, *blue*, *grey*, *light grey*, and *brown*. In each

[2] http://en.wikipedia.org/wiki/K-means_clustering

group, colours are re-ordered by observers concerning with the lightness variants. The agreed clusters (demonstrated in Table 5.1) are embraced and used for the brush colour picker design of System A, as shown by Figure 5.2. System A represents an example of the *discrete-discrete* interface.

TABLE 5.1: Distribution of 256 discrete colours in 12 colour categories.

Colour sets	red	orange	yellow	green	cyan	blue	purple	pink	brown	grey	black	white	bright
256 *k-means* colours	48	0	46	63	0	28	0	0	33	14	0	0	24
256 gradient colours	24	24	24	38	24	24	24	24	24	24	1	1	0

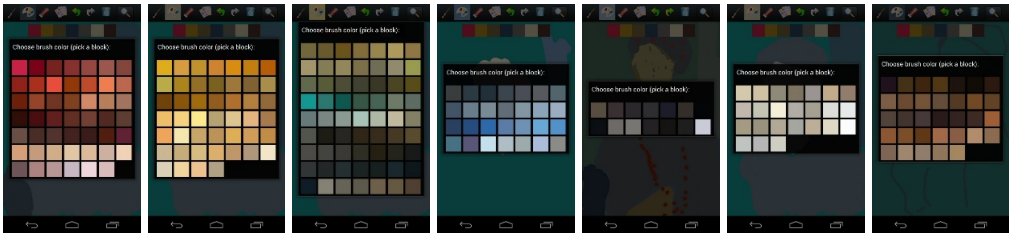


FIGURE 5.2: Screenshots of System A's brush colour palette for portrait painting search in Experiment 1.

As also presented by Table 5.1 are 256 gradient colours consist of the shades from 12 colour categories: 24 different shades of each of seven different *hues*, 38 shades of *green*, 30 shades of *red*, 18 shades of *grey*, and 1 shade of *black* and 1 shade of *white*. The colour categories are widely agreed [9] and they are reasonably separated in colour space. These colour shades of each category are generated via the *w3school*[3] website. 256 gradient colours are built as System C's brush colour picker, with 11 hue colours are first shown as the indexes of colour groups, shown from left to right: *red*, *orange*, *yellow*, *green*, *cyan*, *blue*, *purple*, *pink*, *brown*, *grey*, and *black&white*. when a colour category is triggered, a dialogue of colour shades will be displayed (see Figure 5.3). The background colour picker of System C is a continuous H-SV *bar-square*, which IS discussed in the next section.

[3] http://www.w3schools.com/tags/ref_colorpicker.asp

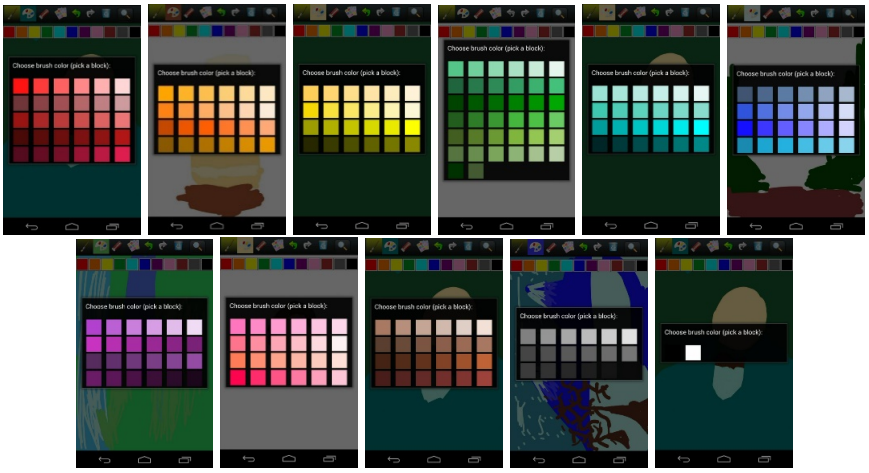


FIGURE 5.3: Screeshots of System C's brush colour palettes (256 gradient colour shades) in Experiment 1.

5.3.3 Continuous Colour Pickers

Two commonly-used continuous colour pickers are implemented. One is the H-SV *bar-square* colour picker, the other is the revised colour *ring* based on the results of the focus group study (please refer to Section 3.3.3).

System C is a prototype for *discrete-continuous* combination. The background colour picker of System C adopts the H-SV *bar-square* representation design. As shown in Figure 5.4, a solid *hue* slider bar stands along a 2D dynamically coloured input square area. In other words, when the user chooses a colour in *hue* slider bar, the square area will accordingly show a variety of shades of the selected *hue*. Both *saturation* and *brightness* of the fixed *hue* are mapped to an XY-plane, on which the *saturation* is related to the Y-axis and the *brightness* is related to the X-axis. Two colour swatches are presented next to an arrow, with the left one showing the current colour and the right one displaying the selected new colour. The 'current $\rightarrow$ new' view would help the user to decide whether the chosen colour is getting closer to the expected colour.

Similarly, the brush colour picker of System B contains several typical display elements: the H-SV *bar-square* with transparency slider, the RGB-alpha sliders,

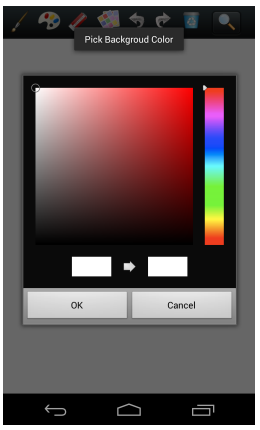


FIGURE 5.4: Screenshot of System C's background colour picker.

and the numeric input textbox (hexadecimal), as shown in Figure 5.5. Users can select a colour from nine colour swatches as starting point, then proceed to finer modification of this base colour by using the H-SV colour square. The nine base colours include *red*, *green*, *yellow*, *orange*, *cyan*, *blue*, *purple*, *pink*, and *brown*, which are corresponding to Berlin and Kay [9], with addition of *cyan*. The new colour swatch is next to the current used colour swatch , with a hint text '*last used color*' or '*new color*' showing on the top of each colour swatch.

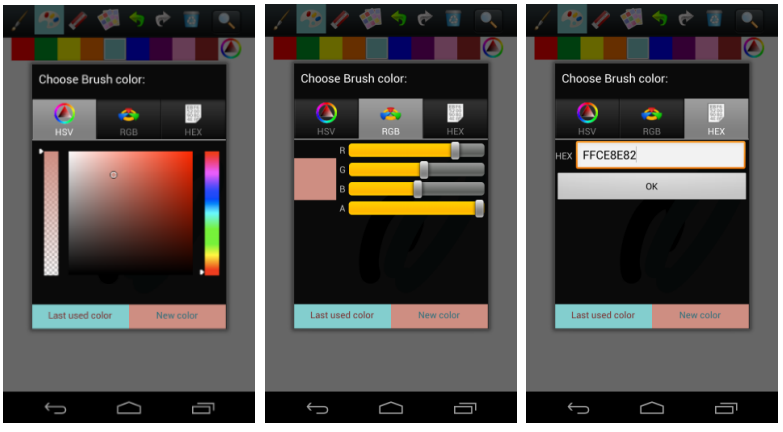


FIGURE 5.5: Screenshots of System B's brush colour picker: H-SV colour square, RGB slider, and Hex value input box.

As mentioned earlier, the revised colour ring will be shown in a pop-up dialogue when the user presses the background colour picker icon, as shown in Figure 5.6.

System B is an example of the *continuous-continuous* colour picker combination interface.



FIGURE 5.6: Screeshot of System B's background colour picker.

5.4 Experimental Design and Procedure

5.4.1 Stimuli

Based on the stimulus selection rationales (see Section 4.3.1) and three experimental task types proposed in Section 4.4.2, the following six stimuli are chosen (shown in Figure 5.7):



(A) *Lady* (B) *Braids* (C) *Red* (D) *Pink* (E) *Green* (F) *White*

FIGURE 5.7: The stimuli with their names used in Experiment 1.

The '*Lady*' was presented to participants continuously under the copying-task condition (T1 task). Due to the fact that the target painting might not be in the returned results, and we expect to examine the user's query refinement, we asked

participants to refine their initial queries of the '*Lady*' at least two times, until the '*Lady*' appears in the first three most-matched result images. The '*Red*', '*Pink*', '*Green*', and '*White*' were used as the stimulus under shortly-present task (T3 task) condition.

The '*Braids*' was used for the target-unspecific task (T2 task). Only a scenario-based description of the '*Braids*' was presented on a sheet of paper given to participants:

> Tom: "*Can anyone help me? I am looking for a painting I saw a year ago at an art museum. My memory is blurry... but it was a portrait of a woman with black hair in two braids, she wore a dark-grey coat. Only her head and shoulders were shown in the middle of a red background. Her skin was of brown colour.*"

5.4.2 Participants

37 participants (17 female, Mean age=27.92, Age range=19-57), 2 left-handed and 35 right-handed, were recruited. All participants were first-time users of our QbD application, and one participant self-indicated having weak sight. Each participant was compensated with five pounds Amazon voucher after completing all tasks of Experiment 1.

5.4.3 Procedure

After signing the consent form, participants were instructed to copy a simple clipart image (see Section 4.3.1.2) on our QbD App until they got familiar with the user interface. Participants were encouraged to draw with wide brush strokes and make sure the colours and proportion of objects accurately. After the practice session, participants were asked to draw queries for the '*Lady*' by using the

three systems (ABC) in random order, with the stimulus continuously shown on reference phone. On one system, participants were allowed to refine their queries at most twice until the '*Lady*' sitting in the first three most-matched results. Then participants were asked to use the next randomly-allocated system with the same procedure. The comparison of the three systems' brush colour pickers and background colour pickers (in the questionnaire **Q3**, see Section 4.5.3.2) would be made when completing T1 tasks on all systems. No time limitation was set for each drawing procedure.

After a short break, participants were given a sheet of paper with T2 task description of the '*Braids*' (see Section 5.4.1) by using System C. Participants had to speak out what they were doing during the query formulation process. There were no limitations on time and number of query attempts. After each search participants were asked to indicate whether the target painting is within the top 21 returned results. When participants found the target painting or decided to give up this task, they were asked to fill out questionnaire **Q4** (see Section 4.5.3.2) and to explain the criteria they employed to judge the search results.

For T3 tasks, participants were informed that they would be shown a portrait painting for 30 seconds, and after that the reference painting would be removed and they were asked to draw out as many details as they could remember using System C. One of the '*Red*', '*Pink*', '*Green*', and '*White*' painting was randomly selected for each participant to perform. Drawing time and query attempts were not limited. For each query, participants were instructed to fill out questionnaire **Q4** (see Section 4.5.3.2) after viewing the returned results.

When all phases were completed, participants were asked to answer several open questions in the rest of post-questionnaire and were debriefed in a short exit interview. The whole procedure of Experiment 1 requires about two hours from each participant.

5.5 Results

452 drawn-queries were produced with relevant drawing procedure videos record. In this section, we only analyse the 305 drawings of the '*Lady*' produced by 36 participants. One female participant dropped out during the experiment and her data was excluded from the analysis. The drawing processes of the '*Lady*' from one male participant were not recorded due to administrative mistakes, hence all measures extracted from video logs from this participant were also excluded. The queries for T2 tasks and T3 tasks will be examined in the next chapter because of a small number of queries produced for T2 tasks and T3 tasks.

The post-questionnaires were completed by 35 participants (45.71% female). We code the user interactions and analysed the search performance (the number of Top21 queries), the query formulation (drawing time, completeness and quality of a query, and drawing pattern), the query modification and participants' subjective rankings of comfort, preference, and overall opinion of our App. We further examine the drawing strategies of using QbD image search application, such as the first interaction with QbD system and the use of the '*brush*', '*eraser*' and '*undo*' tools, and query refinement pattern.

5.5.1 Statistical Analysis of Copying-tasks

Due to the unequal number of refined queries for the '*Lady*' made on each of the three systems, and the content of the refined query only slightly differs from its initial query from an individual participant, only the initial queries on each system are considered to explore the effects of system design on drawing quality and usability. 107 first drawings are available for analysis.

The independent variables are system design, colour picker type, gender, drawing skill and ethnic group. The outcome variables include the time of query formulation, the number of colours and the number of objects depicted in a query, the number of Top21 queries, participants' usability ratings and evaluators' ratings of colour accuracy. For these outcome measures, the Shapiro-Wilk's tests ($p<.001$) and the visual inspection of their histograms show that the outcome variables violate the normal distribution. A Levene's test indicates that unequal variances in our samples (violated the homogeneity of variance) ($p<.05$). Participants were asked to draw the '*Lady*' repeatedly using the three systems, so the response variables are correlated. The *Generalized Estimating Equations* (GEE) model [86] is useful for these correlated outcome variables [60]. Hence, the GEE method is used to examine the relationships between the independent variables and response variables. For the continuous outcome measures, the gamma distribution is used; for the count data (e.g., the number of colours and objects), the negative binomial distribution is used. Only those main effects and the effects of the two-way interactions between the independent variables that are significant on outcome measures will be reported (with a significance level at α=0.05).

5.5.1.1 Time of Drawing a Query

107 data points from 35 participants were used to analyse the effects of the potential factors on the total drawing time. A GEE model analysis indicates that there was a highly significant effect of system design on the time of query formulation (Wald $\chi^2(2)=14.26$, $p<.0001$). Participants spent much **less time** in drawing a query using **System C** (M=435.71s, SE=31.01) than using System B (M=559.71s, SE=35.48) ($p<.0001$) and System A (M=490.87s, SE=43.71), as shown in Figure 5.8.

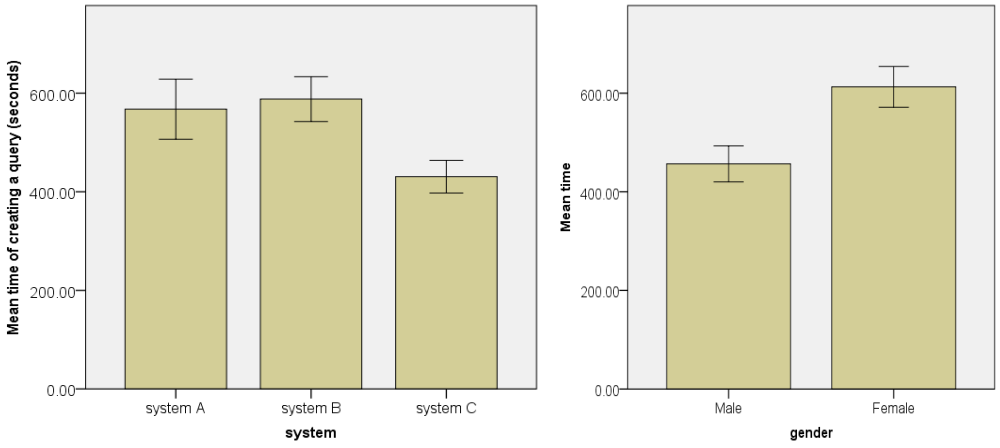


FIGURE 5.8: Main effects of system design (left) and gender (right) on the time of creating a query. Error bars are the standard error of the mean.

Participant's gender also affected the time of drawing a query (Wald $\chi^2(1)$=5.34, p=.021), in that females (M=579.85s, SE=52.41) spent a much longer time in making a query than males (M=420.17s, SE=37.84). Surprisingly, there was no significant effect of participant's drawing skill on the time of drawing a query (Wald $\chi^2(3)$=1.56, p=.668). This non-significant drawing skill effect verifies the participant sampling in our experiment, telling us that if we ignore participant's drawing skill, the drawing time of a copying-task on three systems will be similar to those when considering the effects of drawing skill.

5.5.1.2 Number of Colours Depicted in a Query

A GEE model based on 107 data points reveals that the system design affected the number of colours depicted in a query (Wald $\chi^2(2)$=7.32, p=.026). There was a significant difference among the number of colours depicted in a query produced by using the three systems. Pairwise comparisons indicate that participants drew **more colours** in a query using **System A** (M=11.32, SE=0.87) than those using System B (M=9.34, SE=0.57) (p=.011) and System C (M=9.63, SE=0.61), see Figure 5.9.

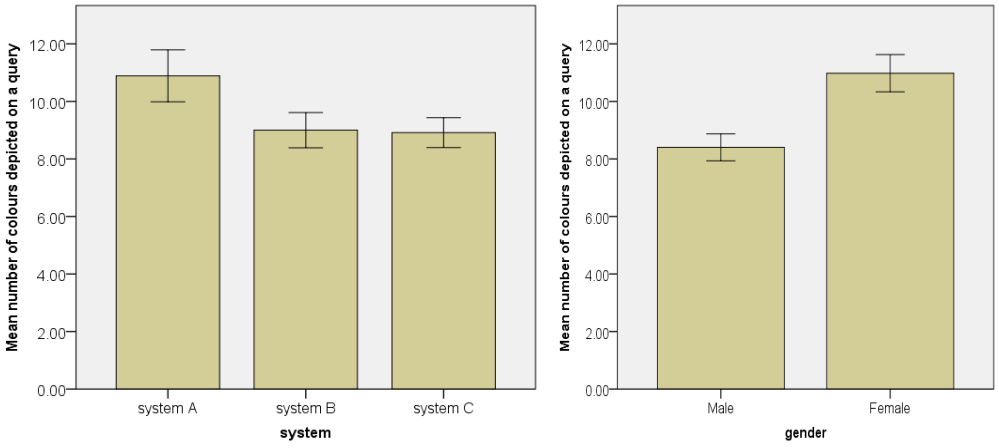


FIGURE 5.9: Main effects of system design (left) and gender (right) on the number of colours depicted in a query. Error bars are the standard error of the mean.

There was also a significant effect of gender on the number of colours (Wald $\chi^2(1)$=8.68, p=.003). **Female** participants (M=11.75, SE=0.99) drew significantly **more colours** for a query than their male counterparts (M=8.61, SE=0.56). No significant effect of participants' drawing skills was found on the number of colours used (Wald $\chi^2(3)$=5.65, p=.130), although we observed that participants with a higher level of drawing skill used more colours than those with lower drawing skill.

5.5.1.3 Number of Objects Depicted in a Query

A GEE model analysis reveals that there was a significant effect of system design in the number of objects depicted in a query (Wald $\chi^2(2)$=10.29, p=.006). Pairwise comparisons indicate that **more objects** were drawn in a query using **System A** (M=13.43, SE=0.89) than those using System B (M=10.80, SE=0.62) (p=.002) and System C (M=11.00, SE=0.57), as shown by Figure 5.10. There was a non-significant difference in the number of objects drawn in a query using System B and System C (p=.709).

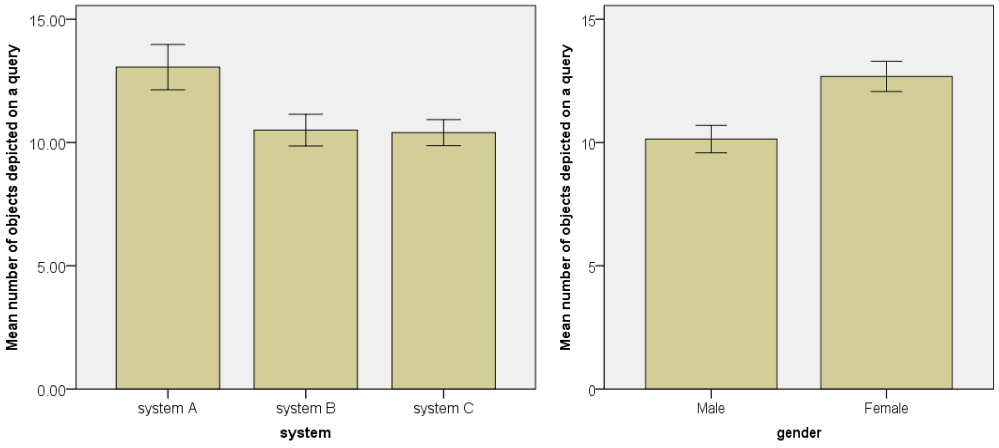


FIGURE 5.10: Main effects of system design (left) and gender (right) on the number of objects depicted in a query. Error bars are the standard error of the mean.

A significant gender effect was found on the number of objects depicted in a query (Wald $\chi^2(1)=10.28$, $p=.001$), in that **females** (M=13.42, SE=0.88) drew **more objects** than males (M=10.18, SE=0.60).

5.5.1.4 Search Performance

The position of the target painting within the returned search results is indicative of search performance. In this experiment, the 21 top-matched images (named Top21 thereafter, Top3 is short for top three results) are considered as retrieval hits. Consequently, the query that hits the target painting in the top N matched results is entitled as the TopN query, such as Top21 query means the query getting the target among the first 21 most-matched images. Due to the stop criteria of tasks (Top3 query), a different number of queries were made on one of the three systems, thus we only analysed search performance of first-attempted queries produced via each system. 108 original queries (36 queries on each of three systems) were examined for search performance measure.

Overall, 65.74% of the first queries made on three systems successfully got the target painting among 21 returned images. It can be seen from Figure 5.11

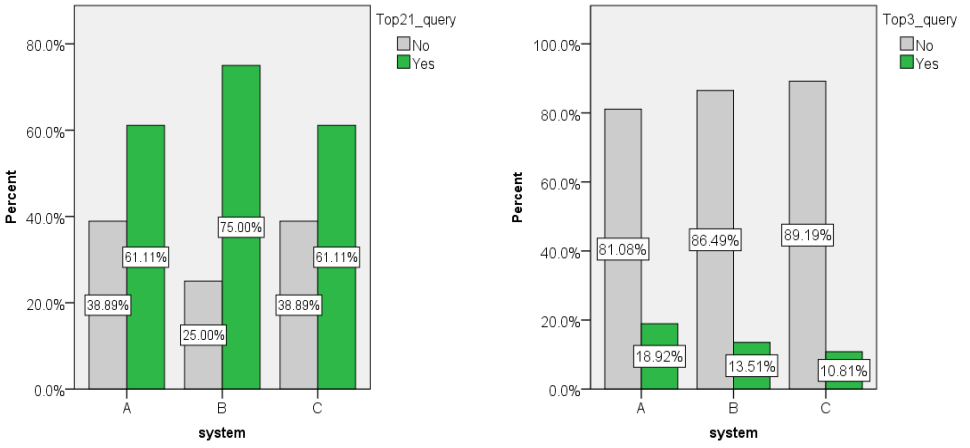


FIGURE 5.11: Proportions of Top21 query and Top3 query made by using each of three systems.

that participants performed best on **System B** among three systems in terms of hitting the target painting in the 21 top matched results. That is, 75% of the first queries made on System B belong to Top21-query. Although the percentages of Top21 queries created on System A and System C are the same (61.11%), participants obtained more queries that hit the '*Lady*' in first three most similar results using System A (18.92%) than using System C (10.81%).

A GEE model reveals that the system design had non-significant effects on the number of Top21 queries (Wald $\chi^2(2)$=2.90, p=.234), and both gender (Wald $\chi^2(1)$=0.44, p=.508) and drawing skill (Wald $\chi^2(3)$=0.73, p=.865) did not affect the number of Top21 queries. The ethnic group was related to the number of Top21 queries (Wald $\chi^2(2)$=8.58, p=.014), with **fewer queries** from **Chinese** participants are Top21-query than those of European participants (p=.002) and participants from other ethnic group (p=.023).

Another GEE model indicates that there was a non-significant effect of system design on the number of Top3 queries (Wald $\chi^2(2)$=1.24, p=.537).

5.5.1.5 Human Ratings on Colour Accuracy

Three evaluators scored the colour accuracy (1=least accurate, 5=most accurate) of the background colour, the foreground colour and the overall colour of a query (as described in Section 4.5.3.5), with comparison to the target painting.

Query's Colour Accuracy 110 triples of query accuracy ratings were assigned by human evaluators, see Figure 5.12.

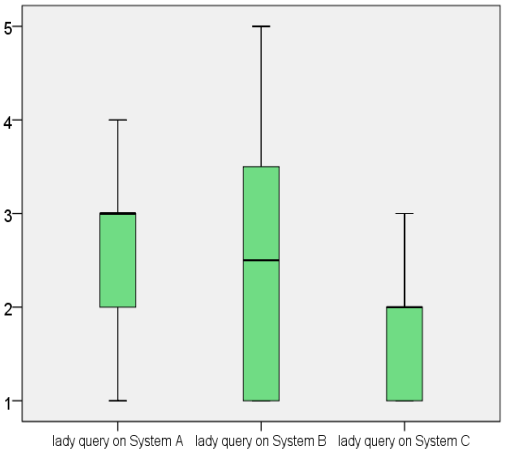


FIGURE 5.12: Boxplot of the query's colour accuracy of the '*Lady*' queries drawn by using System A, B, and C.

The GEE model analysis suggests that system design was significantly associated with the query accuracy ratings (Wald $\chi^2(2)=24.84$, $p<.0001$), with query drawn on **System A** ($p<.0001$) and on **System B** ($p=.028$) was more likely to be rated higher than that on System C, and query accuracy difference between System B and System A was non-significant ($p=.882$).

The GEE model also reveals that **females'** queries were more likely to be rated **higher overall accuracy** than the males' queries (Wald $\chi^2(1)=6.59$, $p=.010$). Queries from **Chinese** participants were more likely to be rated **lower** than those from other participants (Wald $\chi^2(2)=9.38$, $p=.009$). Participant's drawing skill did not affect the query's overall colour accuracy.

Foreground Colour Accuracy 110 foreground colour ratings were collected, see Figure 5.13. A GEE model with multinomial distribution suggests that the colour picker type did not affect the foreground colour accuracy ratings (Wald $\chi^2(1)=0.92$, $p=.337$).

The GEE model also reveals that **females'** queries were more likely to show **higher foreground colour accuracy** than males' queries (Wald $\chi^2(1)=9.18$, $p=.002$). **Chinese** participants' queries were rated much **lower** than other participants' (Wald $\chi^2(2)=12.98$, $p=.002$). Participant's drawing skill did not associate with the foreground colour accuracy rating (Wald $\chi^2(3)=2.42$, $p=.491$).

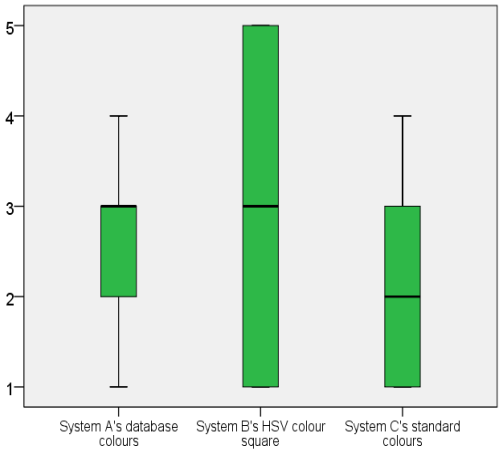


FIGURE 5.13: The boxplot of the foreground colour accuracy ratings of the '*Lady*' queries.

Background Colour Accuracy 109 background colour accuracy ratings were collected, see Figure 5.17 (left). The GEE model reveals that colour picker type was related to background (BG) colour accuracy (Wald $\chi^2(1)=4.77$, $p=.029$), with queries created using **discrete** colour picker were rated **higher background colour accuracy** than those using continuous colour pickers.

The background colour of **females'** queries were more frequently to be rated of higher accuracy score than that of males' queries (Wald $\chi^2(1)=5.55$, $p=.018$).

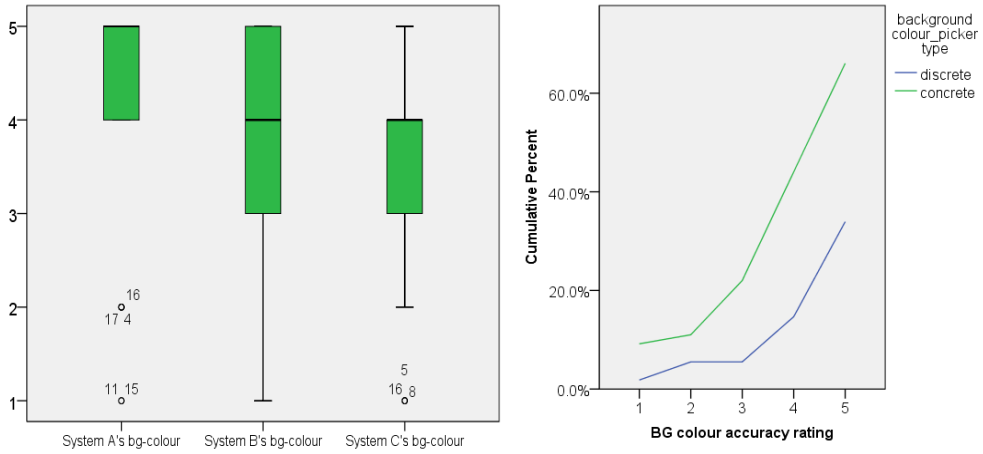


FIGURE 5.14: The boxplot (left) and the cumulative percent (right) of the background colour accuracy ratings of the '*Lady*' queries.

European participants' queries were assigned **higher** background colour accuracy score than Chinese participants' (Wald $\chi^2(1)=7.50$, $p=.023$). Participants' drawing skills did not affect the background colour accuracy of a query.

5.5.1.6 Participants' Usability Choices

Questionnaire **Q3** lists eight usability features: the first two features are relevant to the entire system (including choosing the system costing the least-effort and the least time and also the system costing the most effort and the most time). The rest of the usability features include easy-to-pick a colour, enough colour choices, and preference of both brush colour picker and background (BG) colour picker, respectively. With comparisons of three systems, participants were asked to choose the best system and the worst system for each of the features after completing T1 task (the '*Lady*') on three systems. The resulting choices were coded into binary data (1=selected, 0=not selected).

Table 5.2 summarises the effects of system design on the usability choices. It is found that System B was most frequently selected as the system with the least-effort consuming, and System B's brush colour picker was easier to use than System C's gradient colours. System A's background colour palette with

k-means clustered database colours was less frequently chosen as the easiest-to-use background colour picker compared with the other two continuous BG colour pickers. Participants perceived that System B's brush colour picker provides the largest number of colour choices, while System A's background colour picker offers the fewest number of colours. Even the brush colour picker of System A and System C have the same number of brush colour choices, participants felt that System C's 256 gradient colours supply more colour options than the 256 *k-means* database colours of System A.

TABLE 5.2: Results of GEE models exploring effects of system design on the participants' choices of each usability feature.

Best-worst choice of	Significance level	System Effects
Least-Effort system	**Wald χ^2(2)=9.37, p=.009**	(**B** > C, A)
Most-Effort system	Wald χ^2(2)=1.25, p=.263	(A, C, B)
Least-time system	Wald χ^2(2)=2.68, p=.261	(B, A, C)
Most-time system	Wald χ^2(2)=0.96, p=.619	(C, A, B)
Easiest-to-use brush colour picker	**Wald χ^2(2)=6.50, p=.039**	(**B** > C, A)
Hardest-to-use brush colour picker	Wald χ^2(2)=0.44, p=.801	(C, A, B)
Easiest-to-use BG colour picker	**Wald χ^2(2)=10.11, p=.006**	(B, C) > **A**
Hardest-to-use BG colour picker	Wald χ^2(2)=2.28, p=.320	(A, B, C)
System with most brush colour choices	**Wald χ^2(2)=13.25, p<.0001**	**B** > **C** > **A**
System with fewest brush colour choices	**Wald χ^2(2)=9.19, p<.0001**	(A, C) > **B**
System with most BG colour choices	**Wald χ^2(2)=10.33, p=.006**	(C, B) > **A**
System with fewest BG colour choices	**Wald χ^2(2)=12.99, p=.002**	**A** > (C, B)

Consistent with the effects of system design, Table 5.3 illustrates the effects of colour picker type on the participants' usability choices. Continuous colour pickers, in both brush colour picker group and background colour picker group, were chosen to be easier-to-use than discrete colour pickers. Participants clearly indicated that the continuous colour picker provides much more colour choices than the discrete colour picker, for both brush colour picker and background colour picker.

TABLE 5.3: Results of GEE models exploring effects of colour picker type on the participants' choices of each usability feature (n.s.: not significant).

Best-worst choices of	Significance level	Colour Picker Type Effects
Easiest-to-use brush colour picker	**Wald χ^2(1)=4.77, p=.029**	Continuous > Discrete
Hardest-to-use brush colour picker	Wald χ^2(1)=0.41, p=.524	n.s.
Easiest-to-use BG colour picker	**Wald χ^2(1)=8.21, p=.004**	Continuous > Discrete
Hardest-to-use BG colour picker	Wald χ^2(1)=0.35, p=.557	n.s.
System with most brush colour choices	**Wald χ^2(1)=20.41, p<.0001**	Continuous > Discrete
System with fewest brush colour choices	**Wald χ^2(1)=8.60, p=.003**	Discrete > Continuous
System with most BG colour choices	**Wald χ^2(1)=7.09, p=.008**	Continuous > Discrete
System with fewest BG colour choices	**Wald χ^2(1)=11.49, p=.001**	Discrete > Continuous

5.5.1.7 Preference for Colour Picker

Regarding the preference choice data extracted from questionnaire **Q3**, as indicated in Figure 5.15 that most participants preferred System B's brush colour picker (62.16%) among three brush colour pickers, and System C's brush colour palette with 256 standard colours was chosen as the most disliked brush colour picker (46.88%). The GEE model with binomial distribution reveals that there were significant differences in preference of three brush colour pickers (Wald χ^2(2)=12.46, p=.002). More participants preferred the brush colour picker of System B compared with that of System A (p=.007) to that of System C (p<.0001). The preference difference between System A and System C was not significant (p=.281). Participant characteristics such as gender and drawing skill did not influence their preferences of brush colour picker.

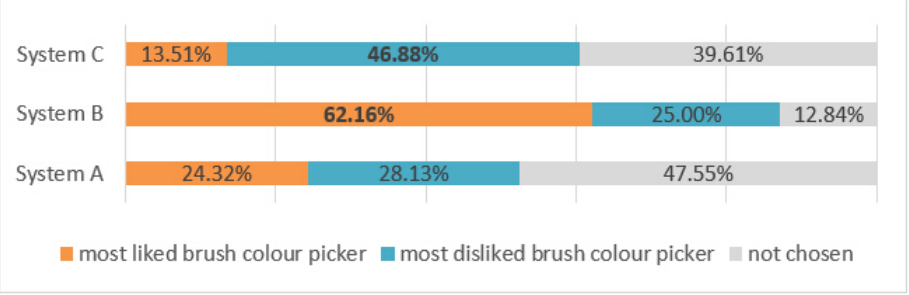


FIGURE 5.15: Distribution of participant's preference of brush colour picker.

With regards to the choice of the most disliked brush colour picker, there were non-significant differences in frequency of the three brush colour pickers being selected (Wald $\chi^2(2)=2.41$, $p=.299$).

In terms of the preference for the background colour picker (see Figure 5.16), 54.29% of participants preferred System B's background colour picker, and System A's 256 *k-means ordered* database colours was selected as the most disliked background colour picker by 53.13% of participants. A GEE model reveals that the participants' choices of the most preferred background colour picker differed significantly (Wald $\chi^2(2)=9.35$, $p=.009$). The pairwise comparisons indicate that the background colour picker of **System B** was much more preferred background colour picker than that of System A ($p<.0001$), and participants preferred the background colour picker of System C than that of System A ($p=.035$). The difference in background colour picker preference between System B and System C was non-significant ($p=.198$).

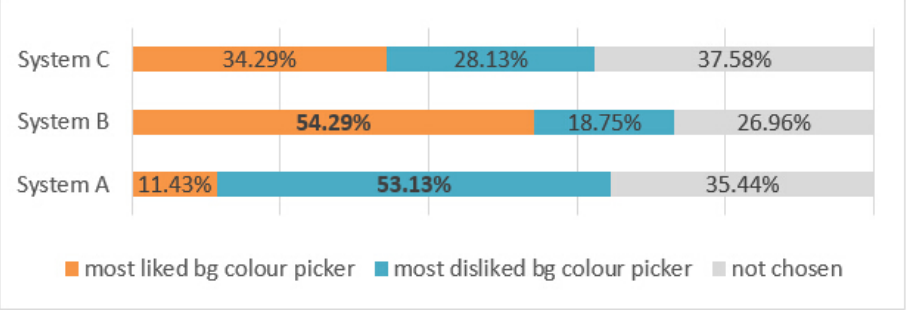


FIGURE 5.16: Distribution of participant's preference of background colour picker.

For the most disliked background colour picker, participants' choices among three background colour pickers were non-significantly different, as suggested by the results of a GEE model (Wald $\chi^2(2)=5.57$, $p=.062$). However, pairwise comparisons show that System A's background colour picker was more frequently selected than System B's ($p=.013$) and System C's ($p=.104$).

TABLE 5.4: Results of GEE models analysing the effects of colour picker type on participants' preference choices (n.s.: not significant).

Preference choice of	Significance level	Effects
Most liked brush colour picker	**Wald χ^2(2)=12.46, p=.002**	**B** > (A, C)
Most disliked brush colour picker	Wald χ^2(2)=2.41, p=.299	n.s.
Most liked BG colour picker	**Wald χ^2(2)=9.35, p=.009**	(B, C) > **A**
Most disliked BG colour picker	Wald χ^2(2)=5.57, p=.062	(A > B, C)
Most liked brush colour picker	**Wald χ^2(1)=10.63, p=.001**	**Continuous** > Discrete
Most disliked brush colour picker	Wald χ^2(1)=1.06, p=.304	n.s.
Most liked brush colour picker	**Wald χ^2(1)=6.43, p=.011**	**Continuous** > Discrete
Most disliked BG colour picker	**Wald χ^2(1)=3.78, p=.052**	(Discrete > Continuous)

Table 5.4 summarises the effects of colour picker design and colour picker type on the participants' preference. Overall, continuous colour pickers are preferred to discrete colour pickers.

5.5.2 Drawing Pattern

We believe that a better understanding of tool-usage behaviour would help to design a more effective system. We only consider the drawing processes of the initial queries of the '*Lady*', resulting in 107 initial queries being collected. The following attributes for the drawing process are analysed: the first interaction with our QbD system, the drawing sequence and the use of brush, eraser and the undo tools.

5.5.2.1 First Interaction with QbD System

TABLE 5.5: Distribution of the first interaction actions for different stimuli.

Stimulus	Choosing BG colour	Outline	Sketch	Block	No. of first queries
Lady	96	4	4	3	107
Braids	25	0	0	3	28
Red	15	0	1	1	17
Pink	12	0	0	0	12
Green	25	0	0	3	17
White	9	0	0	3	9

The first interaction with the system and the use of each tool are discerned by the experimenter by reviewing the video logs. Four typical first interactions with the system are recognised: *choosing background colour*, *drawing overall outline*, *drawing partial sketch*, and *selecting brush colour then creating blocks*. As demonstrated in Table 5.5, most of the queries in the copying task are produced by starting with a background colour, e.g., 96 out of 107 first queries of the '*Lady*' are started with selecting a background colour (three participants selected the background colour as the final step); almost all initial queries in T2 task (the '*Braids*') and T3 tasks are filled by a background colour first, except one query of the '*Red*' started with a sketch and one with a colour block. We also found that the most common second interaction for all tasks is drawing a colour blob. When tracing the foreground of the '*Lady*', 36.49% of participants first drew a contour and then filled colours, which is the way of using a pencil on a paper to draw.

The creation of the background might be affected by the colour picker design. For instance, Participant 34 chose a colour by using a brush colour picker to draw the background because: "*I draw the background (with brush strokes), because (System B's brush colour picker) HSV colour picker is easier than colour wheel (System B's background colour picker).*"

5.5.2.2 Drawing Sequence

We also analyse the sequencing patterns of drawing moves, including *top-down*, *bottom-up*, *left-to-right*, and *right-to-left*. Although obvious *top-down* or *bottom-up* patterns are not found in all portrait drawings, a tendency of *bottom-up* sequences occur in drawings of the *vase* and *flowers* in the '*Lady*' painting (T1 task). 91.67% of first queries for the '*Lady*' are created with the *vase* and *flowers* drawn in a *bottom-up* sequence.

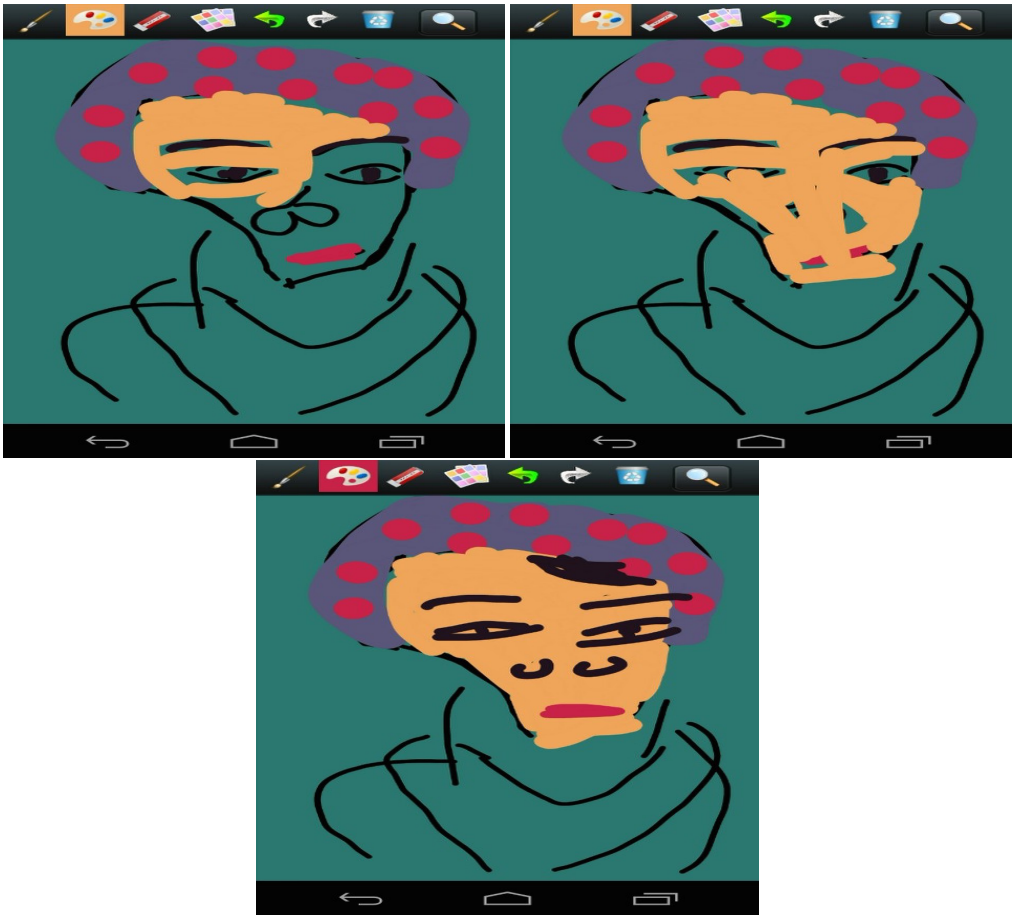


FIGURE 5.17: Example of drawing a contour and filling colours of the '*Lady*' queries.

It is evident that the *left-to-right* sequencing pattern appears in drawings of the target-unspecific (T2 task) and model shortly-present condition (T3 tasks), predominantly for paired objects such as eyes, hairs and arms, which is in line with Goodnow's findings [55]. This might be due to participants' *left-to-right* writing habits learnt from school. Only one of two left-handed participants drew in a *right-to-left* fashion. Participants made fewer *left-to-right* or *right-to-left* sequence when the model is continuously presented during drawing (T1 task). Participants tended to draw in a more systematic way, a *salience-based* way, which is different from with each other due to varied focuses on details.

5.5.2.3 Uses of Tools

It is observed that participants used the brush with default size more often than those with other sizes. Especially in the case of performing the model-absent task (T2 task) and model shortly-present tasks (T3 tasks), most participants only used the brush with default size to complete a query. Participants intuitively picked the '*eraser*' tool to correct mistakes, and they found the '*undo*' function was more efficient afterwards. Overall, the '*undo*' tool was used more often than the '*eraser*'.

Only a few participants explored and used the opacity bar on System B's brush colour picker, particularly people with art and drawing experience. For example, Participant 3 tried to use half-transparent strokes to reconstruct the light depicted in the painting, as he stated: "*It is not easy to draw in terms of representing the light, the material (in the painting) without transparency property.*" Only one participant used the RGB sliders and hex value of System B, and one participant explored the options of System B's brush colour picker but not used actually.

5.5.3 Query Refinement

Regarding the T1 task (copying the '*Lady*'), as required by the task, participants stopped modifying their queries until their queries could find the target painting in top three most-matched results. Therefore, there were a different number of queries produced on each of the three systems during query refinement. Among the first and second query refinement, participants performed best on System B with *continuous-continuous* colour pickers in terms of the number of successful queries (71.11% of refined queries are Top21 queries), followed by System C (59.65%).

The observations of the '*Lady*' drawings and their video logs indicate that participants would engage in query refinement with incremental or decremental

detail modification to make progress, and some participants preferred to draw a new query from scratch, which may be behavioural. We identified 11 types of refinement actions, and the frequency of each action is shown in Table 5.6. Most of the initial queries (85.71%) were refined with at least one or two types of modification actions.

TABLE 5.6: Frequency of 11 types of refinement action for the '*Lady*' task.

Refinement actions	Frequency
Change foreground colour	**50.68%**
Add new object(s)	**32.43%**
Change background colour	**29.05%**
Add object detail(s)	14.86%
Change object(s) position	10.81%
Add background colour	8.11%
Change object(s) size	7.43%
New whole query	6.08%
Remove object details(s)	4.05%
Remove object(s)	3.38%
New foreground	1.35%

It is quite apparent that the most frequently used action of query modification in the '*Lady*' task is **changing foreground colour** (50.68% queries), followed by **adding new object(s)** (32.43%), **changing background colour** (29.05%), and **adding details of an object** (14.86%). This seems to validate that our design of separating the brush colour selection from background colour selection facilitates users to refine their queries. We noted that 11 modified queries were **made from scratch** (by using the '*Trash*' bin to delete the entire drawing firstly) from seven participants, who also modified three queries by re-drawing the entire foreground (using the '*Eraser*' to erase all strokes firstly). It might be the individual habit that caused the modification difference.

For T2 task (the '*Braids*') and T3 tasks (drawing after 30-second viewing), the query refinement is not compulsory. However, 18 participants voluntarily changed their first queries: 11 participants modified queries for the '*Red*', and six changes queries for the '*Green*', and one query for each of the rest of stimuli ('*Braids*', '*Pink*', and '*White*') were refined. The changing background colour

was used most often, followed by changing foreground colour(s). However, due to the small number of samples, the refinement pattern for the T2 tasks and T3 tasks needs to be validated further.

It is worthy to note that four participants, who have knowledge of Content-based Image Retrieval, refined their initial queries based on comparisons between their queries and several top matched images in the search results to decide which part of drawing should be modified. However, most participants, particularly those with limited expertise in CBIR, just tried to guess the misleading areas in the drawing, which often made the refinement an iterative *trial-and-error* procedure.

5.5.4 Search Strategies

Not all search behaviour could be easily explained. The overall search process includes *query formulation*, *result browsing*, *enlarging thumbnail*, and *query modification*. A *maximal repeating pattern* (MRP) analysis [119] shows that the most common pattern is firstly drawing a query, then browsing the thumbnail results, and then going back and forth between the state of modifying drawing and the state of viewing new search results. The vast majority of participants browsed the thumbnail of result images to judge the result relevance, and only two participants visited the large view of thumbnail image.

Even though participants were informed that accurate colour and proportion are more important for creating a successful query, participants were uncertain when to submit their drawing to search and what a good query is. This might be because the task requirement of maximum three attempts limits participants' free exploration with unlimited query modifications. It is noted that 21.62% of participants started with a simple drawing and then refined the query by *trial*

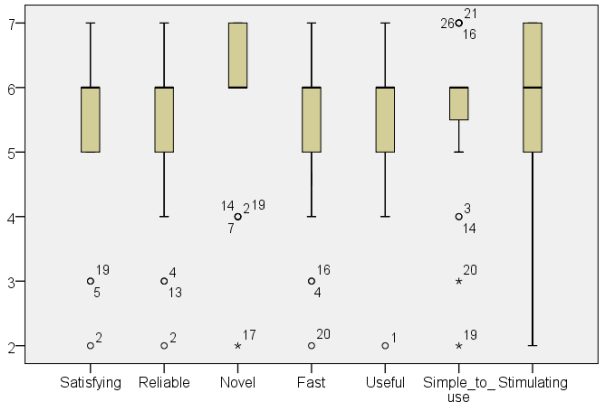


FIGURE 5.18: Participants' ratings on seven attributes of our QbD App.

and error. Three participants who have knowledge of CBIR technology refined their drawings through learning from the returned results.

5.5.5 Opinions and Suggestions on Overall System

It is indicated that most participants rated positively on our system's attributes with respect to the search results (*satisfying*, *reliable*, and *fast*), the App functionality (*novel* and *useful*), and the user interface (*simple-to-use* and *stimulating*), see Figure 5.18. As expressed by Participant 35:"... *I like the search function as well, because it is quite exciting to see if my drawing is accurate enough, i.e., getting a target in the top three (results).*"

The following insights have been extracted from participant's suggestions during the exit-interview, which is helpful to improve the next version of our QbD App. Two participants suggested to have several simulated pen or pencil or brush strokes styles (Participant 33:"...*to increase the types of brush; increase the function of drawing a series of shapes, such as circle, triangle...*") as well as the image effects of a query before submitting to search, such as the *blurred brush* effect and *sepia* colour effect. Participant 35:"...*some photo effects may be needed in portrait drawings, e.g., blur of brush effects...*" Participant 36:"*To*

supply different shapes with different tasks, such as pen tool in Photoshop, and sepia colour effect, which is old-fashion colour..."

Many participants indicated that the functionality of keeping the used colours is necessary, thus the users could find previously used colour. Regarding the discrete colour palette, it would be better to mark the location of the last selected colour swatch, otherwise users may forget the colour group from which they have chosen, especially when picking the same or a closed colour via the colour palette with *k-means clustering* database colours.

5.6 Discussion

In this section, we will discuss some findings from Experiment 1, which offers a guidance for developers to decide using between continuous colour picker and discrete colour picker and their combinations. Many of people have intuition in this decision, e.g., the discrete colour picker might be less time-consuming. However, we seek to provide an empirical advice by systematically analysing the implications of choosing appropriate colour picker for a Query-by-Drawing image search mobile App.

5.6.1 Effects of Colour Picker Type

Two groups of colour pickers are examined in this experiment: one is the brush colour picker group, including the colour palette with *256 human-ordered database colours* (System A), the *V-HS bar-square* (System B), and the colour palette with *256 standard gradient colours* (System C); and the other group is the background colour picker group, including the colour palette with *256 k-means clustering database colours* (System A), the *Colour ring* (System B), and the *V-HS bar-square* (System C).

TABLE 5.7: Summary of colour picker type effects on measures for the '*Lady*' queries (n.s.: not significant).

Measures	Effects	Significance level
Foreground colour accuracy	n.s.	p=.337
Background colour accuracy	**Discrete** > Continuous	***p*=.029**
Easiest-to-use brush colour picker	**Continuous** > Discrete	***p*=.029**
Hardest-to-use brush colour picker	n.s.	p=.524
Easiest-to-use BG colour picker	**Continuous** > Discrete	***p*=.004**
Hardest-to-use BG colour picker	n.s.	p=.557
Most brush colour choices	**Continuous** > Discrete	***p*<.0001**
Fewest brush colour choices	**Discrete** > Continuous	***p*=.003**
Most BG colour choices	**Continuous** > Discrete	***p*=.008**
Fewest BG colour choices	**Discrete** > Continuous	***p*=.001**
Most liked brush colour picker	**Continuous** > Discrete	***p*=.001**
Most disliked brush colour picker	n.s.	p=.304
Most liked BG colour picker	**Continuous** > Discrete	***p*=.011**
Most disliked BG colour picker	(**Discrete** > Continuous)	p=.052

Table 5.7 presents the overall effects of two colour picker types. Among the background colour group, the background colour of a query produced by using the discrete colour picker was rated a higher accuracy score than that using the continuous colour picker. Participants perceived that the continuous colour picker in both brush and background colour picker group is easier to use than the discrete one. Because of the limited number of available colours, it is difficult for participants to select the expected colours. Similar to the 256 gradient colours, it might be the case that none of the 256 database colours perfectly matches a colour in the user's mind, thus he or she must come to a compromise between the available colours and the expected colour. As expressed by Participant 30:"*I can't find 'purple', and I definitely want 'purple'. This (colour) is too dark* purple*, for instance, and that (colour) is too light. The colour I want is just in-between (these two colours).*" Moreover, most participants reported that they were confused by the random display sequence of colours on System A's brush colour picker and background colour palette, even though System A's brush colour palette is manually re-ordered and grouped according to the human perceived brightness. Participants had clear views that the colour choices of the

continuous colour picker are much more than those of the discrete colour picker.

According to participants' preference choices, most participants preferred to use the continuous colour picker for selecting both brush colour and background colour, and disliked the discrete background colour picker. Several reasons for why users like or dislike a colour picker, which are excerpted from participants' transcripts during the exit-interview, are listed as follows:

Participant 34: "*I think it* (the coat colour of the 'Lady') *was between pink and blue, but I do not know how to pick a colour and from which group* (on System C's 256 colours), *that is why I do not like the colour-block palette. While when I use the HSV colour palette, I can compare so many different colours, and I can pick the colour directly! It could also remember which colour I used before.*"

Participant 27: "*I preferred selecting a colour directly rather than using a gradient* (HSV bar-square) *because I felt more inspired that way. It* (colour swatches) *was also easier. The gradient* (continuous colour picker) *was very difficult to use because my fingers are so thick! It was hard to choose colours.*"

Participant 25: "*I preferred choosing from a wide range of colours rather than colour swatches in order to match them perfectly. I did not like the colour options* (colour swatches) *where I could not change the shade.*"

In summary, participants who prefer the continuous colour picker emphasise on its ability of picking the exact colour that cannot be found in the discrete colour picker, while participants who prefer the discrete colour palette focus on the ease-of-use feature that allowed them to select a colour swatch rather than mixing a colour by using a full-choice colour picker. Several participants preferred System B's brush colour picker because its ability to keep the last used colours, as mentioned by Participant 34.

It is worthy to notice that we only explore two types of discrete colour picker, one with random lightness, and one with gradient lightness. It is possible that participants' opinion on discrete colour picker is affected by the K-means background colour picker, because sometimes participants could not find an expected colour easily and quickly.

5.6.2 Effects of System Design

The combinations of continuous colour picker and discrete colour picker are implemented into three systems, with a brush colour picker and a background colour picker (brush-background combination) for each system:

- System A: *discrete-discrete* (256 ordered k-means database colours - 256 k-means database colours)
- System B: *continuous-continuous* (V-HS bar-square - Colour ring)
- System C: *discrete-continuous* (256 gradient colours - V-HS bar-square)

There are measurable differences in how the participants performed on three systems, as shown in Table 5.8. Participants drew more colours and objects in a query by using the *discrete-discrete* colour pickers (System A). The reason might be that discrete colour picker is easier to select a colour to try and hence participants could concentrate on more details in a query. Oppositely, as Participant 35 stated: "*It* (System B's brush colour picker) *took a long time to pick the right colour, so I draw roughly, not like* (drawing on System) *A*."

Participant spent the shortest time in drawing a query on the *discrete-continuous* System C. It is clear that time for drawing foreground accounts for most of total time of drawing a query, and System A's discrete background colour palette takes less time to pick a background colour than System C's background colour

TABLE 5.8: Summary of System effects on measures for the '*Lady*' queries (n.s.: not significant).

Measures	Effects	Significance level
Total time of query formulation	(B, A) > **C**	***p*<.0001**
No. of colours depicted in a query	**A** > (C, B)	***p*=.026**
No. of objects depicted in a query	**A**> (C, B)	***p*=.006**
Number of Top3 queries	n.s.	*p*=.537
Number of Top21 queries	n.s.	*p*=.234
Colour accuracy of a query	(B, A) > **C**	***p*=.010**
Foreground colour accuracy	n.s.	*p*=.020
Background colour accuracy	n.s.	*p*=.020
Least-effort system	(**B** > C, A)	***p*=.009**
Most-effort system	n.s.	*p*=.263
Least-time system	n.s.	*p*=.261
Most-time system	n.s.	*p*=.619

picker, therefore we may deduce that System A's 256 database brush colours cost significantly more time to draw the foreground of a drawing than System C's 256 gradient brush colours. But it might because that more details are drawn when using System A, and thus it takes longer time to complete a query on System A than that on System C.

Although search performance (the number of Top3 queries and Top21 queries) did not differ among the three systems, the queries produced on **System C** were rated the lowest accuracy scores by human evaluators in terms of the foreground colour and query's overall colour. A positive correlation is found between query colour accuracy ratings and foreground colour accuracy ratings, but it raises more concerns that require further exploration, e.g., increasing sample size and more types of stimuli.

More participants perceived that they spent less effort in using the *continuous-continuous* system (System B) than using the *discrete-continuous* (System C), but the choices of the system with the least time-consuming did not vary significantly among the three systems.

Due to the limited time and resource, we have not examined and compare the *continuous-discrete* combination with other three. Because the background colour picker is less frequently used (maybe once or twice for drawing a query) comparing to the brush colour picker, we believe that the performance of a *continuous-discrete* system would be similar to that of a *continuous-continuous* system. Therefore we feel that the findings of our experimental comparisons between three combinations of colour picker type is sufficient as the representative design of QbD image search applications.

5.6.3 Effects of User Characteristics

Participants are divided into groups based on gender, ethnic group and drawing skill. Effects of gender are found on the total time of query formulation, and the number of objects and the number of colours depicted in a query.

Our findings of gender differences are in line with Koppitz [77] that females tend to include more details than males, which could explain that female participants spent significantly longer time in drawing a query than males. Moreover, females' queries are more frequently to be rated higher accuracy scores in terms of the foreground colour, background colour, and overall query accuracy than males'.

Consistent effects of ethnic group are found on both objective and subjective measures of query accuracy, in that *Chinese* participants produced fewer Top21 queries than other participants. Both foreground colour and query's overall colour of a query from Chinese participants were assigned lower accuracy scores than that from other participants. This might be because the stimulus paintings are western oil painting in which Chinese participants are not interested. But we do not know the real reasons for this.

Surprisingly, non-significant effects of drawing skill are found, although it is observed that participants with higher drawing skill tend to be serious on selecting a more accurate colour to draw a query. These non-significant effects verify our participants sampling, telling us that if we ignore participant's drawing skill, the outcome measures would be similar to those considering participant's drawing skill. However, it might be arbitrary, and larger sample size are needed to verify this.

5.6.4 Drawing Process

When drawing a query for all three types of tasks, it is found that participants firstly selected a background colour, followed by drawing colour blobs in the canvas, which is the use pattern expected by thesis author. Some novice users (36.49%) drew their first queries in the same way as using the pencil-on-the-paper, which is making a contour firstly and then filling colours within the contour. With target painting in sight, participants traced the foreground objects in the painting in a systematic sequence (e.g., *saliency*-based); while for target-unspecific task or target-absent task, participants usually drew a query in a *left-to-right* pattern, even one of two left-handed participants drew in a left-to-right fashion. Hence, it is not a good idea to put the tool panel in the left or right side of the canvas. However, most participants seemed to be uncertain when to submit a query to search. During query refinement, participants firstly modified the foreground, e.g., *changing the foreground (object) colour(s)*, or *adding more objects or small details* of existing objects. Changing background colour is also firstly concerned by some participants.

It is worthwhile to note the issue of practice effects in this study since the participants were required to copy one painting and searching and refining for several times consecutively on each of three systems. Participants might lack patience to do the same task repeatedly and that is why we assigned the three systems in

random sequence to the participants. As one left-handed participant indicated: "*copying the portrait lady is unpleasant and annoying. It was boring to copy.*"

5.7 Summary

This chapter mainly examines how colour picker types (continuous and discrete) and their combinations (three combinations of discrete and continuous colour pickers) affect the query performance and usability for performing a copying-task. We also compare the effects of the different colour picker interfaces (three discrete colour palettes, three continuous colour pickers) on the usability and user preference. Meanwhile, an overview of how users draw and search and modify their queries by using our mobile QbD application are also included.

Overall, participants spend the least time in drawing a query by using the *discrete-continuous* system (System C), and more colours and objects are depicted in a query when using the *discrete-discrete* system (System A). The system design has a non-significant influence on the search performance measures. However, based on human evaluators' accuracy ratings, queries produced by using the *discrete-continuous* System C are rated lowest accuracy scores in terms of foreground colour and overall colour of query. This experiment also offers some insights into participants' preference for colour picker and system usability, in that the **continuous** colour picker is **preferred** and participants perceived the **continuous** colour picker is **easier** to use than the discrete colour picker.

In summary, participants thoroughly have positive attitudes on our QbD App and they enjoy the idea of drawing to search. They are satisfied with the general user interfaces, especially the design of separation of background colour picker from the brush colour picker. Participants generally draw a query by starting with selecting a background colour, followed by drawing the colour patches with brushes. For the copying-tasks, foreground objects are generated in a systematic

sequence, while for the target-unspecific or target-absent tasks, a *left-to-right* sequence emerges from participant's drawing process. Participants follow the '*trial and error*' way to search a query and refine a query and search again, etc. Modifying foreground colours, adding objects, and changing background colours are the three most frequently used actions during query refinement. It seems that most participants are not clear when to submit a query and which part of an initial query should be modified for 'better' results.

Chapter 6

Experiment 2 - Task Type, Painting Type and Brush Type

6.1 Introduction

Drawing with the painting in sight (T1 task) is just one scenario of using Query-by-Drawing image search. In most cases, the target painting is not present when conducting a search. Real-life image search can be categorised into target-specific tasks (e.g., previously-seen painting) and target-unspecific tasks (e.g., painting with specific theme). The colour selection procedure might be affected by the task type, because the target-specific task requires a higher similarity of a picked colour to the previously-seen colour, while the correctness of mental colours of some objects are required for the target-unspecific task [145]. Landscape paintings were not included in stimuli in Experiment 1. Furthermore, only one target-absent portrait painting (T2 task) and four target shortly-present portrait paintings (T3 task) were tested by using the *discrete-continuous* system (System C) only. We are interested in validating the findings of Experiment 1

on landscape painting stimuli and more portrait painting stimuli as well as under other types of task conditions.

Real paintings normally do not contain regions of plain colour because of brush features and canvas properties such as roughness and oriented strands. Together with different intensities and colours, 'visual texture' is formed on oil paintings [84]. To make a realistic painting like drawing, an intuitive solution is to provide certain pre-defined texture brushes. For example, Town and Sinclair [128] proposed a CBIR image search prototype for ordinary users to construct a query with texture brushes of 11 semantic categories (see Figure 6.1). Similarly, we have designed and implemented two systems with several groups of texture palettes for portrait painting and landscape painting search respectively. The new texture systems were evaluated and compared to three systems with plain colour brushes (refer to the 'Colour System' in short thereafter) in Experiment 2. The experimental comparisons allow us to examine the effects of task type, painting type, brush type and participant characteristics on task performance, preference and usability of our QbD App. Experiment 2 also explores drawing patterns and system choice for real use scenarios.

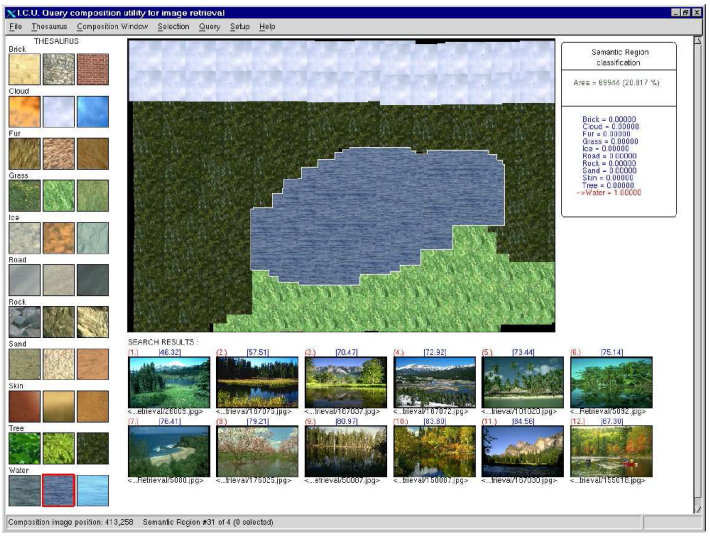


FIGURE 6.1: Screenshot of the query formulation tool taken from Town and Sinclair [128].

This chapter is structured as follows: in Section 6.2 we introduce the purpose of Experiment 2. We describe the database and the refined colour picker interface used in Experiment 2 in Section 6.3 and two systems with texture brushes are briefly introduced in Section 6.4. Section 6.5 presents the experimental design and procedure. In Section 6.6, we first show the statistical analysis results of queries for all copying tasks to validate the findings of Experiment 1 (Section 6.6.1). The effects of task type and painting type on the usability are discussed in Section 6.6.2 and Section 6.6.3, and the effects of brush type and how people use the textural brush are described in Section 6.6.4. Section 6.6.5 introduces several topics in the real task scenario (T2 tasks and T3 tasks) from Experiment 1 and Experiment 2. We also present participants' drawing behaviour when conducting target-absent tasks (Section 6.6.6) and the result relevance judgement strategies used for target-unspecific tasks (Section 6.6.7). A general discussion is given in Section 6.7.

6.2 Aims of Experiment

The primary purpose of Experiment 2 is to examine how the usability of our QbD App is affected by task type, painting type and brush type. Target-specific tasks and target-unspecific tasks are designed to understand how users use the drawing interface and the feasibility of using QbD search paradigm in these scenarios. We would also like to extract hidden patterns in user query formulation behaviour when the user has freedom to choose a system to make queries under real-world circumstances (e.g., drawing natural objects such as grass, sky, and Caucasian skin). Experiment 2 also serves as a supplementary study to Experiment 1. With more stimuli and participants, we aim to verify the findings of Experiment 1. Overall, the purposes of Experiment 2 comprise:

- To validate the findings of Experiment 1 with more stimuli and more task types.
- To explore the effects of task type, painting type and brush type on usability and user preference, if any.
- To examine how different types of tasks would lead the participants to employ one system over another.
- To collect users' performance and use behaviour when using textural brushes with comparison to those of using plain colour brushes.

This experiment will explore the following research questions (RQ):

RQ3 : How do task type, painting type and brush type affect task performance, usability and user preference of a QbD mobile application?

RQ2 : How do users use different user interfaces to draw queries for different types of tasks?

6.3 Colour Picker Interfaces

The search for portrait paintings and that for landscape paintings are separated, but the same user interfaces of drawing tools are used. The users have to indicate their expected painting type before making a query by selecting '*Portrait*' or '*Landscape*' on the welcome screen, as shown in Figure 6.2, to search for a portrait or a landscape painting. Except for the database-related colour pickers (on System A), the continuous colour pickers and 256 gradient colour palette are the same colour pickers used in Experiment 1.

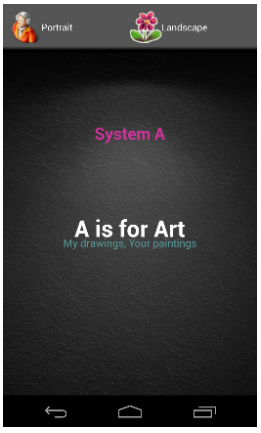


FIGURE 6.2: Welcome screen of our QbD App used in Experiment 2.

6.3.1 Database and Result Displaying

In Experiment 2, the database contains 400 portrait paintings and 400 landscape paintings downloaded from *BBC your paintings*[1]. Portrait painting queries are matched against the portrait painting database, and landscape painting queries are matched against the landscape painting database. The returned results are increased to the 51 top matched images. Therefore, the successful query is defined as the query that returns the target painting in the 51 top matched results (**Top51-query**).

6.3.2 Discrete Colour Pickers

By adopting the same approach of extracting and grouping 256 database colours of System A explored in Experiment 1, different brush colour palettes and background colour palettes are generated from the portrait painting database and landscape painting database individually. In other words, on System A, the 256 *k-means* colours for producing the portrait painting query differ from those for drawing the landscape painting query. The '*check*' sign is put on top of the colour swatch that is last chosen to indicate the previously picked colour.

[1] http://www.bbc.co.uk/arts/yourpaintings/

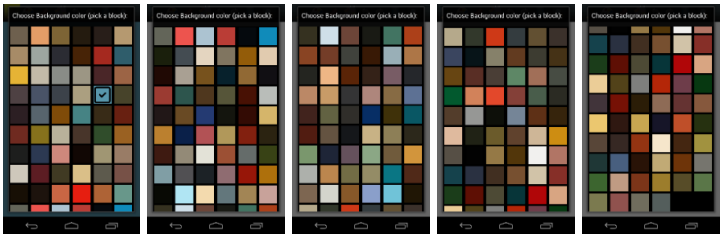


FIGURE 6.3: Screenshots of System A's background colour palette for portrait painting search in Experiment 2.

For the background colour pickers, 256 k-means colours are presented in a dialogue, with six colour boxes in a row, in the order of *k-means* clustering output (please refer to Figure B.2 and Figure B.3 in Appendix B for colour values of k-means '*Portrait*' and '*Landscape*' background colour), as shown in Figure 6.3 for '*Portrait*' background colour palette and Figure 6.4 for '*Landscape*' background colour palette.

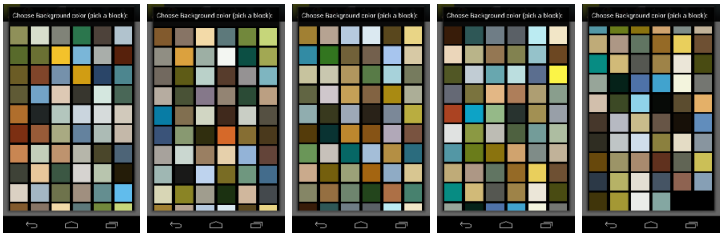


FIGURE 6.4: Screenshots of System A's background colour palette for landscape painting search in Experiment 2.

TABLE 6.1: Distribution of 256 *k-mean* brush colours for portrait painting and landscape painting search in seven colour categories.

Brush colour sets	red	yellow	green	blue	brown	grey	bright
[Portrait] 256 *k-means* colours	54	38	50	30	41	31	12
[Landscape] 256 *k-means* colours	40	40	70	45	30	18	13

For each brush colour picker, there are seven colour groups. In each group, six colour swatches are shown in a row. However, the number of colours (see Table 6.1), the sequence of colours, and the colour values are completely different for portrait painting and landscape painting search, see Figure 6.5 showing the brush colour picker for portrait painting and Figure 6.6 showing the brush colour picker for landscape painting. As illustrated by Table 6.1, the landscape brush

colour picker has more shades of *green* and *blue*, while the portrait brush colour picker has more shades of *red* and *brown*.

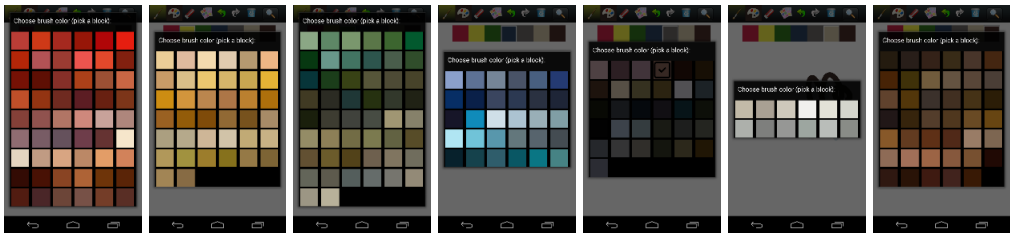


FIGURE 6.5: Screenshots of System A's brush colour palette for portrait painting search in Experiment 2.

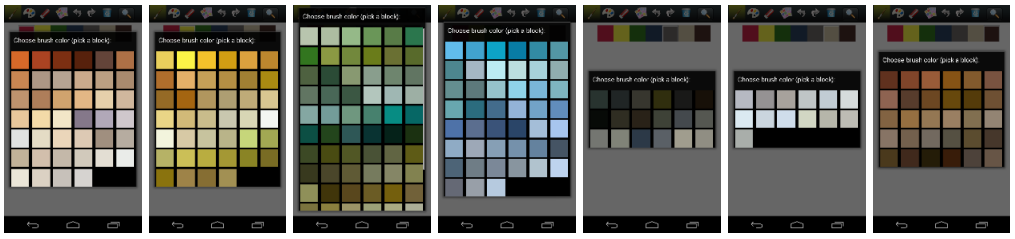


FIGURE 6.6: Screenshots of System A's brush colour palette for landscape painting search in Experiment 2.

6.4 Textural Brush System Design

The systems with texture brushes for portrait and landscape painting search were built and entitled '*Portrait*' system and '*Landscape*' system, respectively. The texture systems allow the user to construct queries by using a menu-dialogue based focal colours and semantic texture palettes, e.g., from a selection of textures of sky, tree, and dress, etc. We anticipate that the texture brush would simplify a user's drawing procedure by avoiding the outlining objects. It is expected that users would use the semantic meanings of these brush textures for making a query. The image retrieval algorithm FMIQ [69] and the same system architecture as in the QbD system with plain colour pickers (see Chapter 3) are adopted.

6.4.1 General User Interfaces

The basic drawing tools include: the '*brush*' with five different sizes, the '*eraser* with five different sizes, '*start new drawing*', and the '*search*' button. The colour palette contains 12 colours which are 11 basic colours proposed by Berlin and Kay [9] and an additional *cyan* colour; and eight groups of texture palettes are sampled from the portrait painting database and the landscape painting database respectively. The user interface was modified based on Sue Smith's work on *tutsplus* website[2]: the textural pattern is repeated throughout the area on which the user draws. And the size of texture does not change with the changing of brush size, rather it would be repeated seamlessly throughout the canvas wherever the user starts with a drawing, as shown in Figure 6.7 (right).

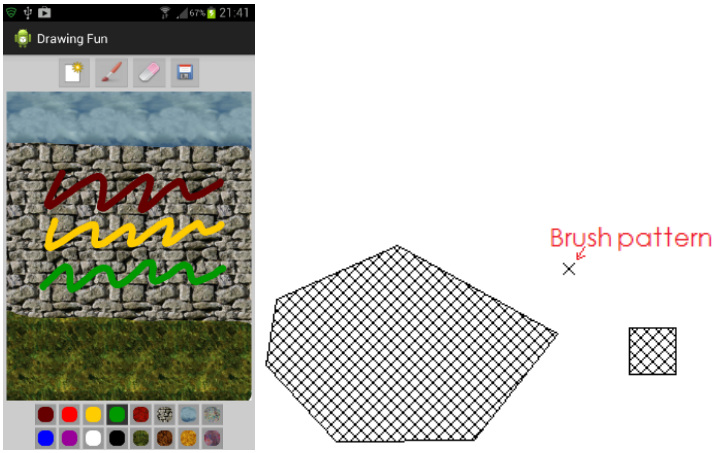


FIGURE 6.7: Screenshot of prototype of System Texture (left) and the repetition mechanism of texture brush (right) (taken from *tutsplus* website).

The author picked the following texture patterns for both portrait and landscape painting by observing the frequency of texture appearing within database images.

[2] http://code.tutsplus.com/tutorials/android-sdk-drawing-with-pattern-fill

6.4.2 Texture Palette of '*Portrait*'

Eight portrait texture palettes are shown from left to right, top-down in Figure 6.8.

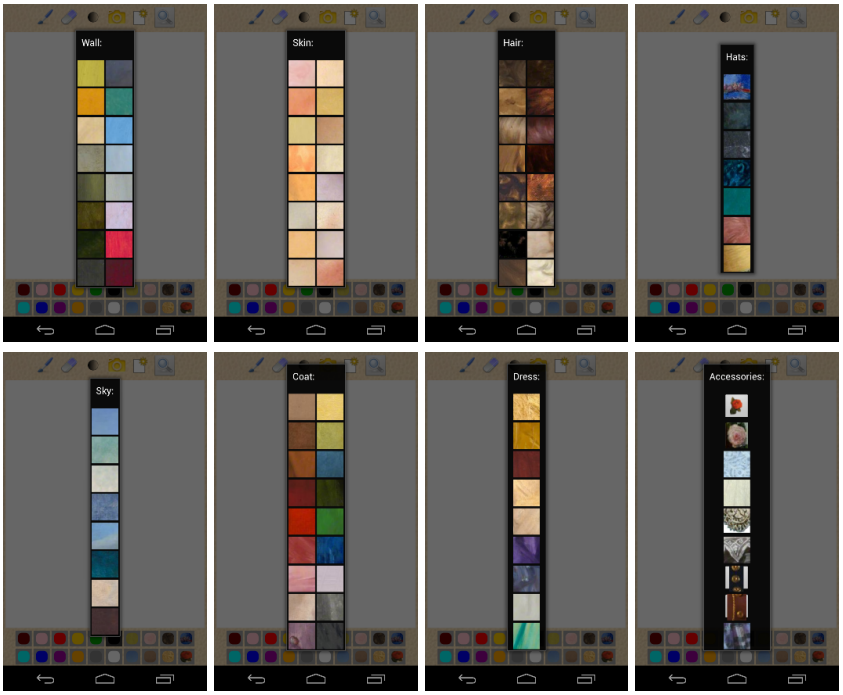


FIGURE 6.8: Screenshots of texture palettes of Wall, Skin, Hair, Hats, Sky, Coat, Dress, Accessories in system '*Portrait*'.

6.4.3 Texture Palette of '*Landscape*'

Eight groups of texture palettes used for landscape painting search are shown from left to right and top-down in Figure 6.9.

6.5 Experimental Design and Procedure

The task conditions include drawing query with a very brief description of target painting (T2 tasks, see Appendix A for verbal description) and drawing painting

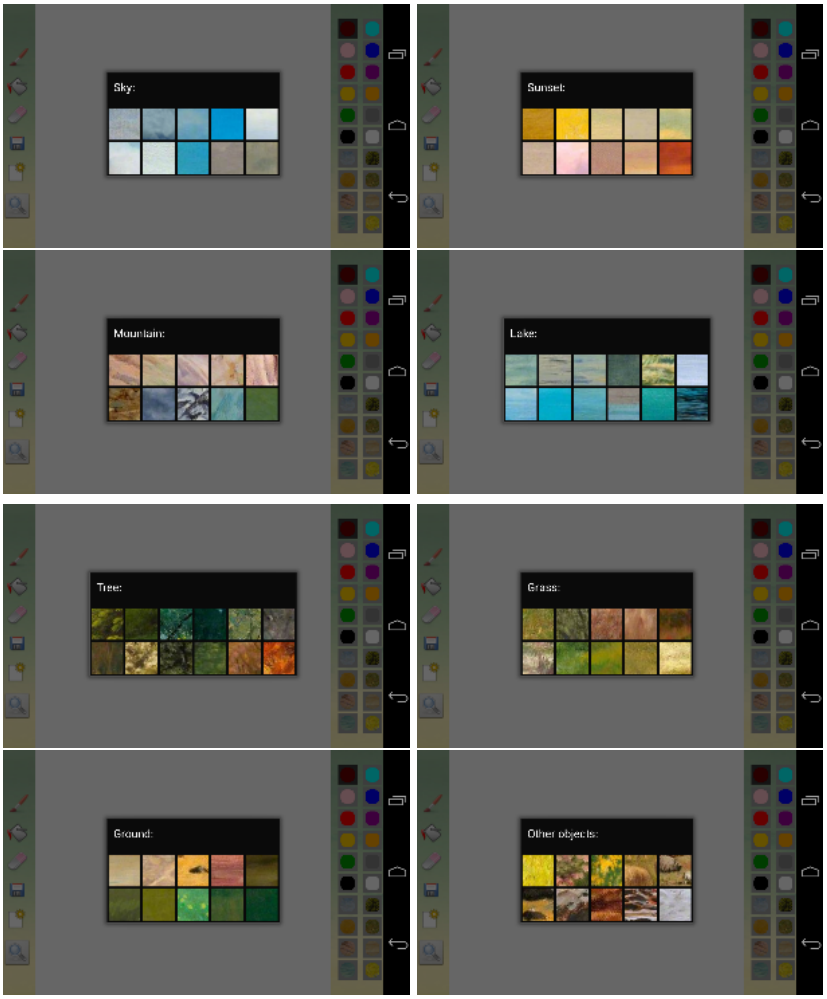


FIGURE 6.9: Screenshots of texture palettes of the Sky, Sunset, Mountain, Lake, Tree, Grass, Ground, and Other objects in system '*Landscape*'.

immediately after 30-second viewing (T3 tasks). In order to restrict potential variances of the participants' subjective drawings, the T1 tasks (copying the stimuli in sight) are also included to provide a basis of comparison between systems. All three types of tasks involve both portrait paintings and landscape paintings.

It is ideal to test all tasks with a within-subjects design. That is, participants draw and search for each painting from all tasks on each of the four systems. However, the participants were found bored to repeat drawing the same things on three systems during Experiment 1. Moreover, each drawing and rating conducting on

one system is expected to take about eight minutes, so that each stimulus on four systems would last approximately 32 minutes, and this would require more than three hours for each participant to draw and search six stimuli on four systems, with addition of sessions for filling the pre-task questionnaire and involving exit-interview. Hence, due to practical limitation, the within-subjects design for T1 tasks (one portrait painting and one landscape painting) in Experiment 2, in which each participant will be asked to trace the continuously-presented stimuli to evaluate and compare the usability of three colour picker systems separately as well as a texture system comparatively. A between-subjects design is used under T2 task and T3 task conditions.

6.5.1 Stimuli

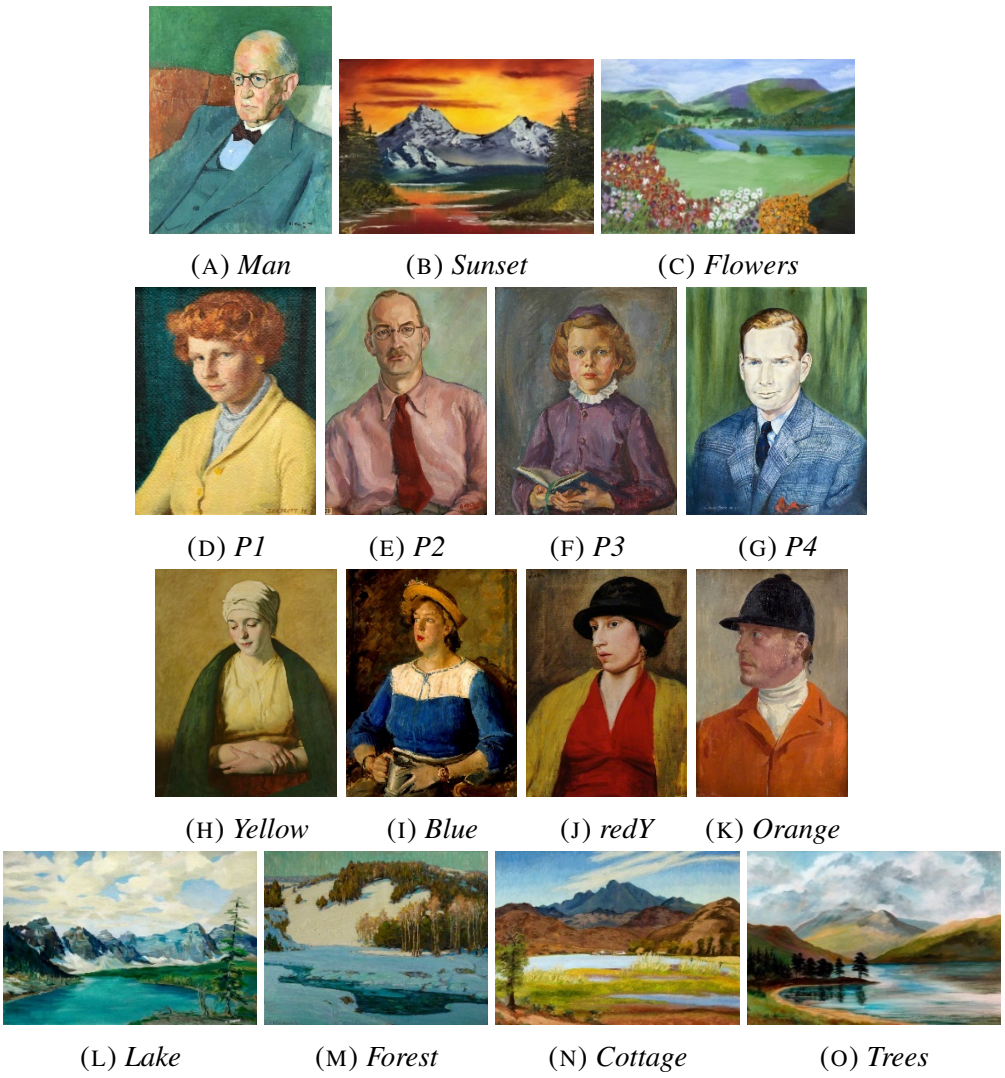
There are 15 stimuli (see Figure 6.10) and 4 target-unspecific seasonal scene used in Experiment 2.

6.5.2 Participants

38 participants (23 male, 15 females, age range=17-29) with three from the Experiment 1 were recruited. All participants indicated having normal colour vision. Demographic data about the participants, including the previous experience of drawing, search engine and colour selection, etc., was also collected through the pre-task questionnaire.

6.5.3 Procedure

Participants were given consent form and pre-questionnaire before the experiment began. The training phase is identical to that of Experiment 1, with participants being asked to draw a query of simple clipart image (see Section 4.3.1.2)



(A) *Man* (B) *Sunset* (C) *Flowers*

(D) *P1* (E) *P2* (F) *P3* (G) *P4*

(H) *Yellow* (I) *Blue* (J) *redY* (K) *Orange*

(L) *Lake* (M) *Forest* (N) *Cottage* (O) *Trees*

FIGURE 6.10: The stimuli with names used in Experiment 2. (A)-(C) are used for T1 tasks, (D)-(G) are the target painting for T2-portrait tasks, (H)-(O) are used for T3 tasks.

using the QbD App until they got familiar with the system functionalities. This experiment is one-by-one with one participant and one experimenter. All drawing sessions took place in a quiet and adequately lit room located in the university. The tasks were randomised across task condition and painting item for each participant. Participants were told to take their time in drawing a query, and neither query attempt nor time limit was imposed for each task.

The first phase of the experiment is similar to that of Experiment 1. Participants

were requested to first draw a query for one portrait painting '*Man*' by using four systems (System A, B, C and Texture) in random order and then search for one landscape painting ('*Sunset*' or '*Flowers*') by drawing on each of four system also in random order. After completing each query on the four systems, the participants were asked to compare the brush colour picker and background colour picker of system A, B, C by voting in the questionnaire **Q3** (see Section 4.5.3.2). They were also required to indicate which kind of brush (plain colour vs. texture) was easier to use and which one they preferred to use when completing all copying-tasks. The refinement for each query was not compulsory.

After a short break, the second phase involving T2 tasks started. For portrait painting search, each participant was assigned task descriptions in random order. For each query, participants could choose any one of the four systems to draw on, even using the same system repeatedly. To ensure that all participants clearly understood these written descriptions, we asked if they had any questions. Once any questions were answered, participant commenced drawing the query. For the four seasonal landscape painting searches, the participants were informed that they had to include at least, but not limited to, five objects in each query: sky, mountain, tree, river, and land, and they had to use all four systems with one seasonal query drawn on one system. For each query, after participants pointed out the relevant image(s) in the results, they were asked to rate the level of agreement of statements in the questionnaire **Q4** (see Section 4.5.3.2).

The third phase of Experiment 2 involved conducting T3 tasks. Each participant was requested to complete the searches for two portrait paintings and two landscape paintings in random order. Participants were allowed to choose any one of the four systems to produce each query. After reviewing the search results of each query, participants were asked to complete the questionnaire **Q4** same as in phase two.

6.6 Results

To make data analysis more reliable, data collected from Experiment 1 and Experiment 2 are examined, with a total of 837 data points of 72 participants' (41 Males, 31 Females) searching for 26 paintings. The results discussed in the following subsections were analysed by a series of *Generalized Estimating Equation* (GEE) models.

Similar to the statistical analysis employed in Experiment 1 (Section 5.5.1), the dependent measures include the task performance measures (the time taken to complete a query, the number of successful queries, the completeness and accuracy of a query), and the scales regarding system usability and user satisfaction of each colour picker. As for the search performance measure, due to the experimental setting - only the first 21 most matched images are returned in Experiment 1, while the first 51 top matches are shown as the search results in Experiment 2. It is reasonable to examine the number of Top21 queries through all queries from two experiments, while the number of Top51 queries should be only counted among queries from Experiment 2.

The independent variables contain gender, ethnic group, drawing skill, task type, painting type, brush type and the colour system design.

We firstly examined the queries for copying tasks (T1 tasks) conducted on three colour systems (ABC) to test and verify the findings about effects of colour picker type and system design from Experiment 1. Then another series of GEE models were used to investigate the effects of task type and painting type on the outcome measures by exploring all queries made during Experiment 1 and Experiment 2, followed by comparisons between colour systems and texture systems. The user behaviour (e.g., drawing pattern and result relevance judgement criteria) emerging in performing target-absent tasks (T2 tasks and T3 tasks) and the effects of ethnic group on these tasks were also specially investigated.

6.6.1 Effects of Colour Picker Type and Combinations

6.6.1.1 Statistical Analysis of Copying-tasks on Colour Systems

There were totally 293 observations of 71 participants performing T1 tasks (two portrait paintings and two landscape paintings) on three colour systems (see Section 5.3): the *discrete-discrete* system A, the *continuous-continuous* system B, and the *discrete-continuous* system C. One participant is excluded because of the lack of video logs.

Table 6.2 demonstrates the effects of system design on the outcome measures of copying tasks (T1 tasks) from Experiment 1 and Experiment 2. It is found that participants depicted more colours and objects in a query when using the *discrete-discrete* system (System A), and spent less time in drawing a query by using the *discrete-continuous* system (System C) than the other two systems.

Overall, 56% of 300 queries for copying-tasks (T1 tasks) are the Top21-query, and 68.75% of 192 queries for T1 tasks produced in Experiment 2 belong to the Top51-query. Figure 6.11 shows the distributions of the successful query among queries produced on each of three colour systems. It is observed that more queries drawn via System B are Top21 queries (59%), and more queries produced on System C belong to Top51 queries (71.44%). However, the GEE model reveals that system design had a non-significant effect on both search performance measures. Human evaluators have assigned a much higher colour accuracy score to the query made on System A than that on System C (p=.004) and this difference between a query made on System A and that on System C was non-significant (p=.064).

From participants' best-worst choices in the questionnaire **Q3**, it is found that System A's background (BG) colour picker was more frequently to be selected as the hardest-to-use BG colour picker and the most disliked BG colour picker, and System B was more often to be chosen as the favourite brush colour picker

TABLE 6.2: Results of GEE models for analysing system effects on all copying tasks from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	System Effects
Total time of query formulation	**Wald χ^2(2)=43.08, p<.0001**	(B, A) > **C**
No. of colours depicted in a query	**Wald χ^2(2)=6.47, p=.039**	**A** > (B, C)
No. of objects depicted in a query	**Wald χ^2(2)=13.66, p=.001**	**A** > (C, B)
Number of Top21 query	Wald χ^2(2)=1.56, p=.481	(B, C, A)
Number of Top51 query	Wald χ^2(2)=1.25, p=.535	(A, C, B)
Colour accuracy of a query	**Wald χ^2(2)=10.37, p=.006**	(**A** > C, B)
Foreground colour accuracy	Wald χ^2(2)=5.36, p=.068	n.s.
Background colour accuracy	Wald χ^2(2)=4.32, p=.115	n.s.
Least-Effort system	Wald χ^2(2)=0.56, p=.756	(B, C, A)
Most-Effort system	Wald χ^2(2)=5.59, p=.101	(B, C, A)
Least-time system	Wald χ^2(2)=1.43, p=.490	(C, A, B)
Most-time system	Wald χ^2(2)=3.41, p=.181	(B, A, C)
Easiest-to-use brush colour picker	Wald χ^2(2)=0.58, p=.750	(A, B, C)
Hardest-to-use brush colour picker	Wald χ^2(2)=3.84, p=.147	(B, A, C)
Easiest-to-use BG colour picker	**Wald χ^2(2)=10.57, p=.005**	(C, B) > **A**
Hardest-to-use BG colour picker	**Wald χ^2(2)=11.00, p=.031**	**A** > (B, C)
System with most brush colour choices	**Wald χ^2(2)=39.83, p<.0001**	**B** > (A, C)
System with fewest brush colour choices	**Wald χ^2(2)=26.53, p<.0001**	(A, C) > **B**
System with most BG colour choices	**Wald χ^2(2)=21.24, p<.0001**	(C, B) > **A**
System with fewest BG colour choices	**Wald χ^2(2)=48.40, p<.0001**	**A** > (C, B)
Most liked brush colour picker	**Wald χ^2(2)=7.75, p=.021**	**B** > (A, C)
Most disliked brush colour picker	Wald χ^2(2)=0.40, p=.818	(B, C, A)
Most liked BG colour picker	**Wald χ^2(2)=22.51, p<.0001**	(C, B) > **A**
Most disliked BG colour picker	**Wald χ^2(2)=24.43, p<.0001**	**A** > (B, C)

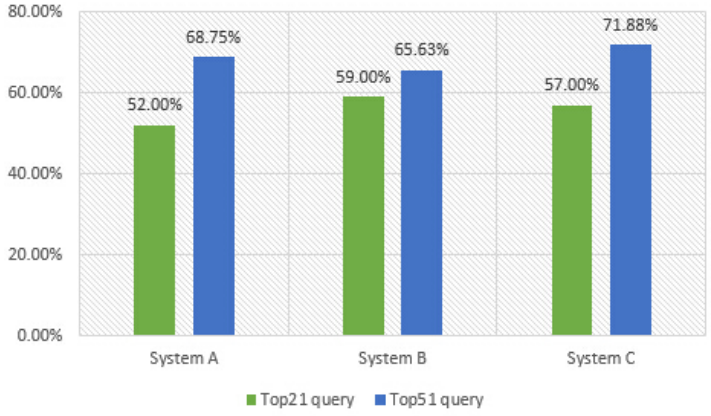


FIGURE 6.11: Percentage of Top21-query and Top51-query made on each system for the copying-tasks in Ex1 & Ex2.

than other two discrete brush colour pickers. Participants had consensus that System B's brush colour picker provides the largest number of colour choices among three brush colour pickers and System A's background colour picker offers the fewest colours among three background colour pickers. It can be inferred that participants perceived more colour choices on the continuous colour picker while a fewer number of colours on the discrete colour picker. This finding is consistent with the effects of colour picker type (as presented in Table 6.3), in that participants selected the continuous colour picker as the colour picker with most colour choices and the discrete colour picker as the one with fewest colour choices. For the two continuous colour pickers - colour wheel and colour square there were no differences in the choice of colour picker preferred and easier-to-use background colour picker.

Table 6.3 also reveals that the continuous colour picker was chosen to be the hardest-to-use brush colour picker but also the easiest-to-use background colour picker, and the discrete colour picker was selected as the hardest-to-use background colour picker. Overall, the **continuous** colour picker in both brush colour picker group and background colour picker group was rated to be the **most liked** colour picker, while the discrete colour picker was chosen to be the most disliked background colour picker.

6.6.1.2 Statistical Analysis of All Tasks on Colour Systems

A total of 684 queries for three types of tasks (26 stimuli) were made on three colour systems (ABC) by 72 participants (41 Males, 31 Females). Table 6.4 summarises the effects of system design on task performance measures. Participants spent more time in drawing a query when using the *continuous-continuous* System B than using others. More colours and objects were depicted in a query by using the *discrete-discrete* system than using other systems.

TABLE 6.3: Results of GEE models for colour picker type effects on all copying tasks from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	Effects
Foreground colour accuracy	Wald $\chi^2(1)$=0.19, p=.666	n.s.
Background colour accuracy	Wald $\chi^2(1)$=0.41, p=.521	n.s.
Easiest-to-use brush colour picker	Wald $\chi^2(1)$=0.29, p=.864	n.s.
Hardest-to-use brush colour picker	Wald $\chi^2(1)$=3.44, p=.064	n.s.
Easiest-to-use BG colour picker	**Wald $\chi^2(1)$=9.66, p=.002**	Continuous > Discrete
Hardest-to-use BG colour picker	**Wald $\chi^2(1)$=7.93, p=.005**	Discrete > Continuous
Most brush colour choices	**Wald $\chi^2(1)$=39.19, p<.0001**	Continuous > Discrete
Fewest brush colour choices	**Wald $\chi^2(1)$=17.53, p<.0001**	Discrete > Continuous
Most BG colour choices	**Wald $\chi^2(1)$=17.52, p<.0001**	Continuous > Discrete
Fewest BG colour choices	**Wald $\chi^2(1)$=39.17, p<.0001**	Discrete > Continuous
Most liked brush colour picker	**Wald $\chi^2(1)$=4.43, p=.035**	Continuous > Discrete
Most disliked brush colour picker	Wald $\chi^2(1)$=0.16, p=.898	n.s.
Most liked BG colour picker	**Wald $\chi^2(1)$=18.69, p<.0001**	Continuous > Discrete
Most disliked BG colour picker	**Wald $\chi^2(1)$=22.50, p<.0001**	Discrete > Continuous

TABLE 6.4: Results of GEE models for analysing system effects on all queries from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	System Effects
Total time of query formulation	**Wald $\chi^2(2)$=13.31, p=.001**	**B** > (A, C)
No. of colours depicted in a query	Wald $\chi^2(2)$=5.29, p=.071	(A > C, B)
No. of objects depicted in a query	**Wald $\chi^2(2)$=6.26, p=.045**	(**A** > B, C)
Number of Top21 query	Wald $\chi^2(2)$=1.56, p=.481	(B, C, A)
Number of Top51 query	Wald $\chi^2(2)$=5.00, p=.082	(A, C, B)
Colour accuracy of a query	**Wald $\chi^2(2)$=15.12, p=.001**	(C, A) > **B**
Foreground colour accuracy	Wald $\chi^2(2)$=0.28, p=.868	n.s.
Background colour accuracy	**Wald $\chi^2(2)$=7.84, p=.020**	(C, A) > **B**

Although the system design had non-significant effects on search performance measures, human evaluators' colour accuracy ratings differed among queries made using the three systems. The background colour accuracy of a query produced using System C's *V-HS colour square* was rated higher score than a query made by using the other two background colour pickers. System A's background colour picker with *k-means clustered* database colours was less frequently to be chosen as the easiest-to-use background colour picker compared with the other two continuous background colour pickers. Participants perceived that System

B's brush colour picker provides the largest number of colour choices and System A's background colour picker offers the fewest colours.

Table 6.5 shows that the colour picker type had non-significant effects on human evaluators' accuracy ratings of foreground colour and background colour.

TABLE 6.5: Results of GEE models for analysing colour picker type effects on all tasks from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	Effects
Foreground colour accuracy	Wald $\chi^2(1)$=0.11, p=.736	n.s.
Background colour accuracy	Wald $\chi^2(1)$=2.52, p=.113	n.s.

6.6.2 Effects of Task Type

The effects of task type are summarised in Table 6.6, which shows that task type highly significantly affected all task performance measurements. Participants spent the longest time in drawing a query for copying-task (T1 task), followed by query drawn from the description (T2 task), and query made after 30-second viewing (T3 task) took the shortest time to complete. Many more colours and objects were depicted in a query when the stimulus is continuously present (T1 task) than that when the stimulus is absent (T2 task and T3 task).

TABLE 6.6: Summary of task type effects on measures for all queries from Ex1 & Ex2.

Measures	Significance level	Effects
Total time of query formulation	**Wald χ^2(2)=247.15, p<.0001**	**T1** > T2 > **T3**
No. of colours depicted in a query	**Wald χ^2(2)=189.73, p<.0001**	**T1** > (T3, T2)
No. of objects depicted in a query	**Wald χ^2(2)=204.68, p<.0001**	**T1** > (T3, T2)
Number of Top21 queries	**Wald χ^2(2)=53.94, *p<.0001**	*(T2, T1) > **T3**
Number of Top51 queries	**Wald χ^2(2)=53.39, *p<.0001**	***T2** > T1 > **T3**
Colour accuracy of a query	**Wald χ^2(2)=109.64, p<.0001**	**T2** > (T1, T3)
Foreground colour accuracy	**Wald χ^2(2)=85.23, p<.0001**	**T2** > (T3, T1)
Background colour accuracy	**Wald χ^2(2)=24.65, p<.0001**	(T1, T2) > **T3**

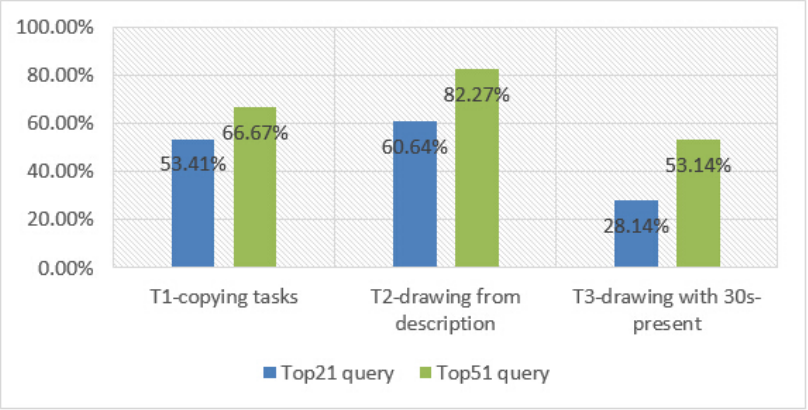


FIGURE 6.12: Percentage of Top21-query and Top51-query among queries for three types of tasks in Ex1 & Ex2.

It is observed that 60.64% of queries for target-unspecific tasks (**T2 tasks**) are Top21-query, followed by 53.41% of queries for copying-tasks, and the fewest number of queries for stimuli 30-second present tasks (T3 tasks) belong to Top21-query. The distributions of Top51 queries among the three task types have similar trends as those of Top21 queries, as shown in Figure 6.12. The results of GEE models verify this observation.

Direct effects of task type on search performance indicate that participants performed **worst** for **T3 tasks** in terms of the number of Top21 queries and the number of Top51 queries among three task types, and participants produced **most Top51 queries** for **target-unspecific tasks** (T2 tasks) than other two types of tasks. The number of Top21 queries for T1 tasks was non-significantly different from that for T2 tasks. However, the painting type interacts with task type. In other words, the painting type modifies the effect of task type on search performance, with the abovementioned effects of task type only occurred during the landscape painting search. For portrait painting search, more queries drawn from the description (T2 tasks) are Top21 queries than those for copying-tasks (T1 tasks), while the difference in the number of Top51 queries between T1 tasks and T3 tasks was non-significant.

Similarly, both foreground colour and overall colour of a query drawn from the

description (T2 task) were assigned a higher accuracy score than those of query drawn from memory (T3 task) and even than that drawn with the stimulus in sight (T1 task). Human evaluators rated the query drawn after 30-sec viewing (T3 task) the lowest background colour accuracy score contrasted to the queries for the other two types of tasks.

6.6.3 Effects of Painting Type - Portrait vs. Landscape

Table 6.7 demonstrates the effect of painting type. A query for portrait painting search took participant more time to complete than that for landscape painting, however, participants drew more colours and objects on a landscape painting query than a portrait one.

The direct effects of painting type on search performance were only found on the number of Top51 queries, in that more landscape painting queries belong to Top51-query than portrait queries, but the interaction between task type and painting type influences this effect. Only for copying-tasks (T1 tasks) and target-unspecific tasks (T2 tasks), more queries for landscape painting search are successful Top51-query than portrait painting search; while for target 30-second present tasks (T3 tasks), participants got more Top51 queries for portrait painting search than those for landscape painting. The painting type difference in the number of Top21 queries was non-significant under all types of task condition.

6.6.4 Effects of Brush Type - Plain colour vs. Texture

The statistical analysis results (see Table 6.8) indicate that System Texture facilitated participants in reducing the time of drawing a query, and fewer number of objects and colours were drawn for a query when using System Texture than

TABLE 6.7: Summary of painting type effects on measures for all queries from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	Effects
Total time of query formulation	**Wald χ^2(1)=5.12, p=.024**	Portrait > Landscape
No. of colours depicted in a query	**Wald χ^2(1)=27.05, p<.0001**	Landscape > Portrait
No. of objects depicted in a query	**Wald χ^2(1)=5.48, p=.019**	Landscape > Portrait
Number of Top21 queries	Wald χ^2(1)=0.17, p=.678	n.s.
Number of Top51 queries	**Wald χ^2(1)=4.98, *p=.026**	*Landscape > Portrait
Colour accuracy of a query	**Wald χ^2(2)=15.85, p<.0001**	Landscape > Portrait
Foreground colour accuracy	**Wald χ^2(2)=85.23, p<.0001**	Landscape > Portrait
Background colour accuracy	**Wald χ^2(1)=4.85, p=.028**	Landscape > Portrait

that using system with colour pickers, it might because of the relative limited choices of brush textures compared to colour options of the colour picker.

TABLE 6.8: Summary of brush type effects on measures for all queries from Ex1 & Ex2 (n.s.: not significant).

Measures	Significance level	Effects
Total time of query formulation	**Wald χ^2(1)=70.55, p<.0001**	Colour > Texture
No. of colours depicted in a query	**Wald χ^2(1)=22.53, p<.0001**	Colour > Texture
No. of objects depicted in a query	**Wald χ^2(1)=17.43, p<.0001**	Colour > Texture
Number of Top21 queries	Wald χ^2(1)=2.90, p=.088	n.s.
Number of Top51 queries	**Wald χ^2(1)=5.30, p=.021**	Colour > Texture
Colour accuracy of a query	**Wald χ^2(1)=6.85, p=.009**	Texture > Colour
Foreground colour accuracy	**Wald χ^2(1)=8.98, p=.003**	Texture > Colour
Background colour accuracy	Wald χ^2(1)=0.66, p=.797	n.s.
Preferred brush type	Wald χ^2(1)=3.58, p=.058	(Texture, Colour)
Easier-to-use brush type	**Wald χ^2(1)=5.55, p=.018**	Texture > Colour

It is found that search performance of the plain colour systems is better than System Texture only in the number of Top51 queries, and when considering the first 21 most matched images, the search performance between using plain colour and textural brushes was not great. According to human evaluators perceived accuracy ratings, queries produced using texture brushes were more accurate than those made using plain colour brushes.

After completing all T1 tasks, participants were asked to select the easiest-to-use brush type and the most preferred brush type with a comparison between

systems with plain colour brush and System Texture. There are 84 responses on the preference of brush type and 70 responses on the easiest-to-use brush. Overall, 59.52% of preference choices and 64.29% of easier-to-use responses are the textural brush. System Texture was preferred for drawing a landscape query (see Figure 6.13) and textural brush was also selected as easier-to-use brush compared to the plain colour brush for both portrait and landscape painting search. For example, Participant 60 indicated that "*I like to use the pattern brush, which makes my drawing real.* "

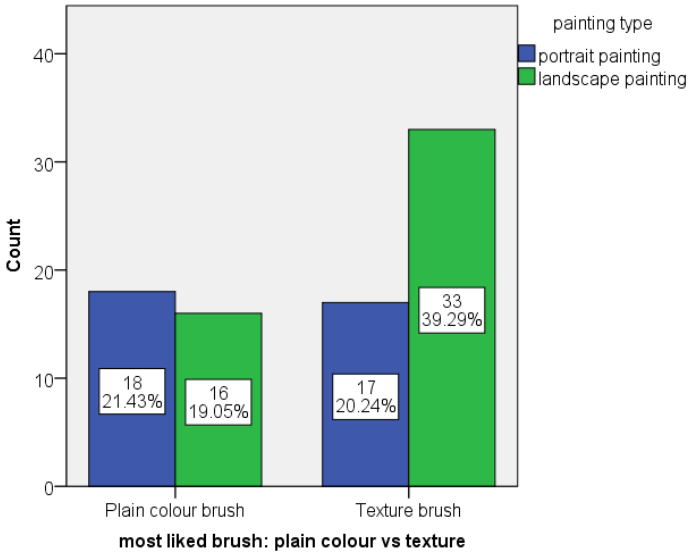


FIGURE 6.13: Distribution of participant brush type preference with respect to painting type.

6.6.5 Real-life Use Scenario - T2 tasks and T3 tasks

6.6.5.1 Ethnic Group Effects

Because the stimuli are sampled from Western oil paintings and the painting description is written in English, we are interested to see the effects of ethnic group on the query formulation, and the accuracy of query in terms of search performance and human accuracy ratings. Based on queries for T2 tasks only, queries

for T3 tasks only, and queries for T2 tasks and T3 tasks, a set of GEE models with same outcome variables and independent variables are computed. Here we only focus on and report the significant effects of ethnic group on queries for T2 tasks and T3 tasks.

487 queries of T2 tasks and T3 tasks from Experiment 1 and Experiment 2 were produced by 70 participants from 20 countries (two participants quit the experiment after completing the copying tasks due to experiment-unrelated reasons), as shown in Figure 6.14. It is observed that most of the queries (180) for target-absent tasks are produced by Chinese participants, followed by English participants (121). Hence the ethnic group variable is organised into three categories: Chinese, European (including English, French, Germany, Spanish, and Poland), and Others ethnic group (including Arabic, Bengali, Egypt, Indonesia, Korean, Kurdish, Malay, Russian, Singapore, Slovene, Thai, Turkish, and Vietnamese).

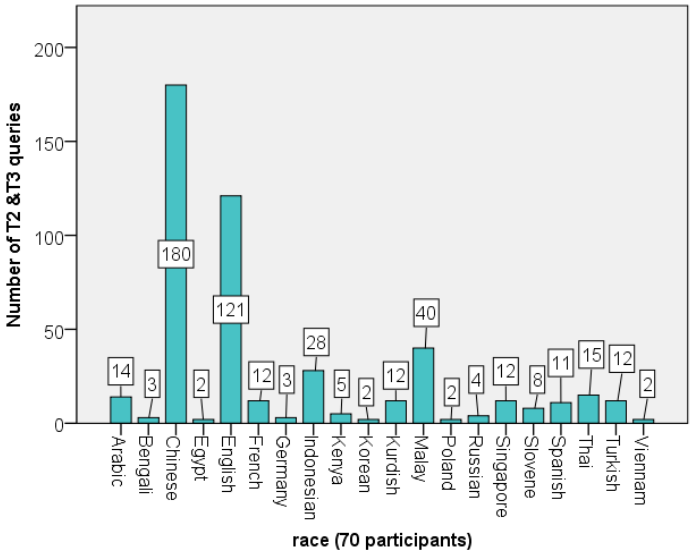


FIGURE 6.14: Distributions of queries made by participants from each of 20 countries.

The effects of ethnic group were found on the drawing time and completeness of a query, as well as the accuracy ratings (as shown in Table 6.9). It shows that for the tasks of searching by drawing from a description (T2 tasks), European participants' queries were rated with higher foreground colour accuracy than

Chinese participants' queries. Regarding queries drawn after 30-second viewing (T3 tasks), queries from European participants were ranked more accurate than those from Others ethnic group, and participants from Others ethnic group spent the shortest time and drew fewer colours and objects in a query compared with European and Chinese participants. Taking the queries for T2 tasks and T3 tasks together, it is found that European participants drew more colours than participants from Others ethnic group.

TABLE 6.9: Significant ethnic group effects on measures for T2 queries, T3 queries, and T2 & T3 queries from Ex1 & Ex2.

Measures	Significance level	Effects
Measure on all T2 queries: Foreground colour accuracy	**Wald χ^2(2)=6.69, p=.035**	European $>$ Chinese
Measure on all T3 queries: Total time of query formulation	**Wald χ^2(1)=11.48, p=.003**	(Chinese, European) $>$ Others
Measure on all T2 & T3 queries: No. of colours depicted in a query	**Wald χ^2(2)=8.56, p=.012**	European $>$ Others

Although we observed that there were differences in search performance of queries among three ethnic groups for each type of task, with participants from Others ethnic group producing more Top51 queries than European and Chinese participants, and European participants made most Top21 queries for T2 tasks, and Chinese participants produced most Top21 queries for T3 tasks, as shown in Figure 6.15. However, the statistical analysis of GEE models reveals that the differences in both the number of Top21 queries and the number of Top51 queries were non-significant between different ethnic groups.

Drawing skill only affected the number of objects depicted in a query for T2 tasks or T3 tasks, that is, participants with higher drawing skill would draw more objects and details than those with lower drawing skill.

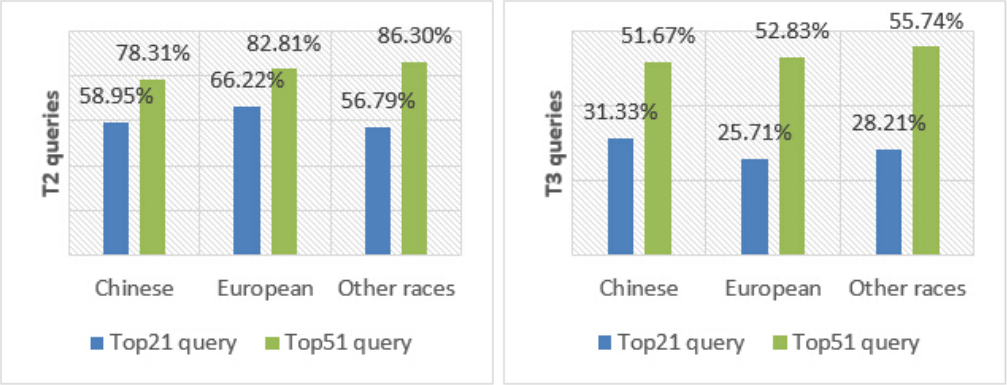


FIGURE 6.15: Distributions of queries made by participants from each of 20 countries.

6.6.5.2 Statement Agreement Ratings

After completing individual search of each T2 task and T3 task, participants were asked to rate on a five-point scale ranging from Strongly Disagree (score of 1) to Strongly Agree (score of 5) the extent to which they agreed with five statements described below. 269 responses on each statement are collected and the resulting scores are formed into ordinal data, and the Generalized Estimating Equations (GEE, only concerning main effects of independent variables) models are used to analyse participants' Likert-scale responses to each statement, concerning with the potential effects of system design, task type, painting type, gender and drawing skill.

Statement 1 : "*I had a mental image of the target painting.*"

When drawing a query from the description (T2 tasks), participants had no idea what the target painting looks, and participants just drew from their imagery. For drawing query after 30-second viewing, participants might preserve the visual memory for previously-seen painting. Overall, most of the queries (62.45% of ranking are assigned score 4, 15.61% are rated score 5) were produced from participant's mental image, see Figure 6.16 (left). The results of a GEE model reveal that drawing skill had a highly significant effect on participants' ratings of

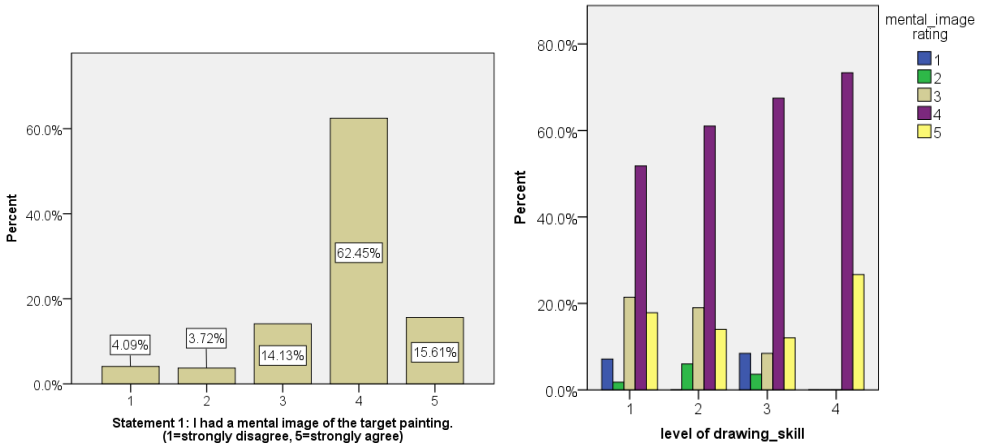


FIGURE 6.16: Distribution of participants' agreement ratings on Statement 1 (left) and the effects of drawing skill (right).

their ability to formulate a mental image of a target painting (Wald $\chi^2(3)=23.20$, $p<.0001$), while the system used or task type or gender have non-significant effects on the mental image scores. As shown in Figure 6.16 (right), participants with higher drawing skill (self-reported score of 4, M=4.32) rated higher mental image scores than those with drawing skill scores of 1 (M=3.76, $p<.0001$), those with a score of 2 (M=3.85, $p=.011$), and those with a score of 3 (M=3.79, $p=.004$).

Statement 2 : "*My drawing was an accurate representation of the painting(s) I had in mind.*"

This statement indicates how precise a participant's drawing represents his or her mental image. Most queries (46.47% of responses are the score of 4, 8.18% are the score of 5) were perceived as accurate representations of participants' search intent, see Figure 6.17. A GEE model analysis suggests that the potential factors, such as system design, task type, painting type, gender and drawing skill, had non-significant effects on participants perceived accuracy ratings.

Statement 3 : "*I am happy with the painting(s) I found.*"

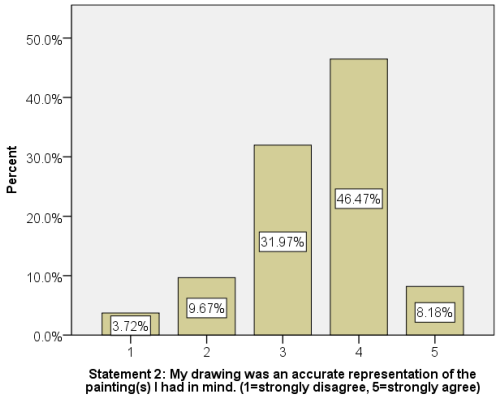


FIGURE 6.17: Distribution of participants' agreement ratings on Statement 2.

It is observed that most responses (41.79% of ratings are allocated score 4, 20.15% are ranked score 5) indicated that participants were happy with their search results, see Figure 6.18. The GEE model analysis reveals that there was a significant effect of task type on participants' ratings of their satisfaction at the results (Wald $\chi^2(1)$=21.94, p<.0001). It seems that participants were more satisfied with the search results of T2 tasks (M=3.63) than those of T3 tasks (M=2.75).

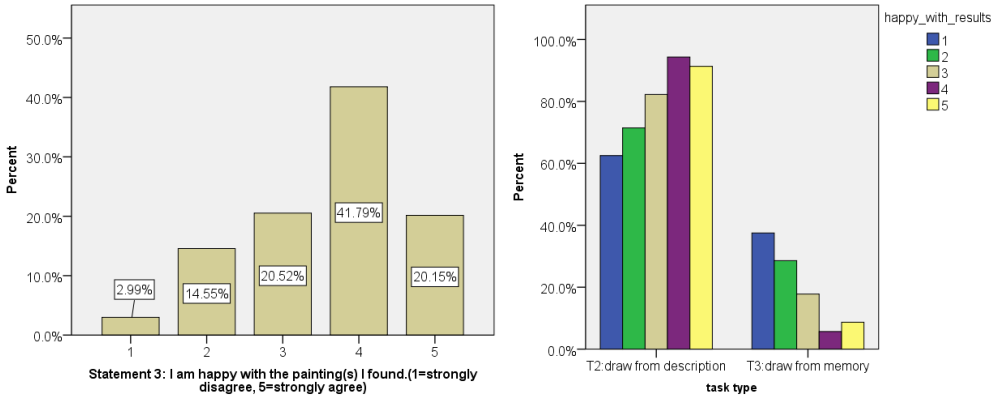


FIGURE 6.18: Distribution of participants' agreement ratings on Statement 3 (left) and the effects of task type (right).

Statement 4 "*I believe that I have seen all possible paintings that satisfy my requirement.*"

A GEE model with main effects of system design, painting type and task type indicates that there was a significant effect of task type (Wald $\chi^2(1)$=15.11, p<.0001) on participants' ratings. Participants believed that they had seen more relevant paintings when performing T2 tasks (M=3.37) than T3 tasks (M=2.46), see Figure 6.19.

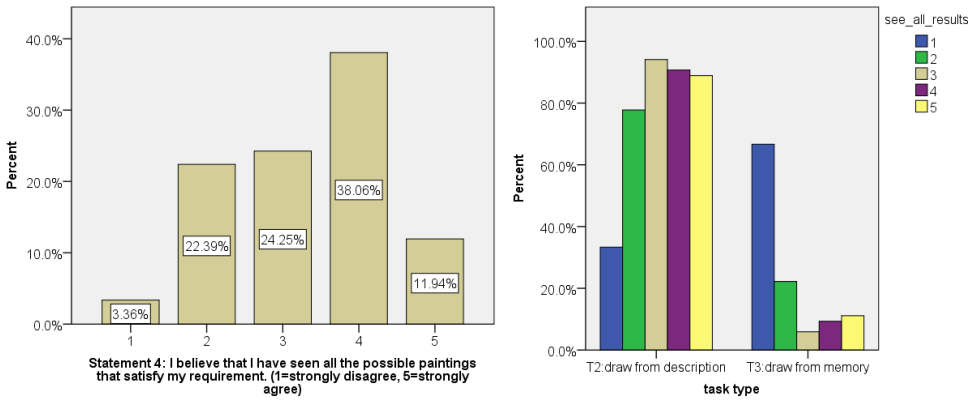


FIGURE 6.19: Distribution of participants' agreement ratings on Statement 4 (left) and the effects of task type (right).

Statement 5 "*I have successfully completed the task.*"

Participants perceived that most queries were successful (49.07% of responses are ranked score 4, 20.15% are ranked score 5), see Figure 6.20. A GEE model analysis reveals that task type significantly affected participants' ratings on task successfulness (Wald $\chi^2(1)$=45.67, p<.0001). Participants believed that they performed more successfully on T2 tasks (M=3.72) than on T3 tasks (M=2.50).

6.6.5.3 System Choices and Task Type

In order to simulate how users may typically use our QbD App in a non-experimental situation, the T2 tasks (drawing from painting description) and T3 tasks (drawing immediately after 30-second viewing stimuli) are designed. Participants

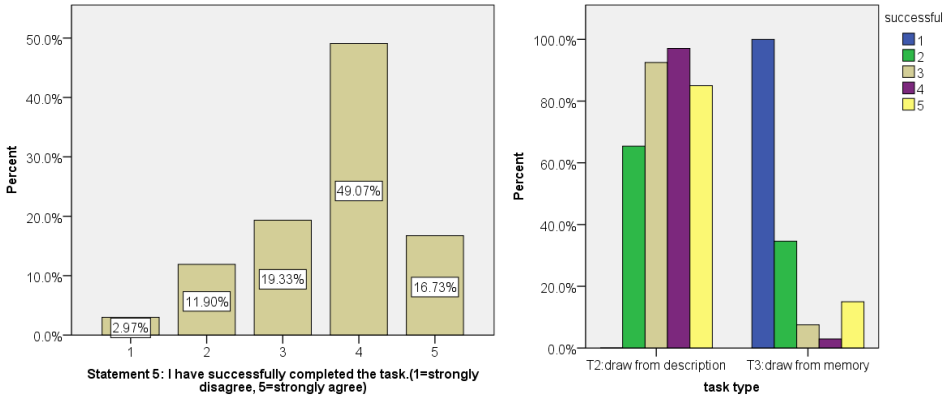


FIGURE 6.20: Distribution of participants' agreement ratings on Statement 5 (left) and the effects of task type (right).

were able to choose from three colour systems and one texture system to perform each task and list the reasons for their system choices.

For T2-portrait tasks and all T3 tasks, 28 participants were asked to use any one of four systems (System A, B, C and Texture) to complete each query, and one system could be repeatedly used. Therefore an unequal number of queries were produced via each of four available systems, and the number of times that system being chosen is an implication of participants' preference of colour picker given this specific scenario.

For T2-landscape tasks - the seasonal landscape painting search, 24 participants were requested to use four systems with one seasonal landscape query produced on one of four systems separately.

T2-Portrait tasks and T3 tasks 238 queries for T2-portrait tasks and T3 tasks were produced by 28 participants. It is observed that System C is the most frequently used system, with 46.22% of queries drawn using System C, followed by 21.85% of queries made via System Texture, see Figure 6.21.

A GEE model analysis reveals that the participants' choices of systems were significantly different among the four systems (Wald $\chi^2(3)$=13.51, p=.004), in

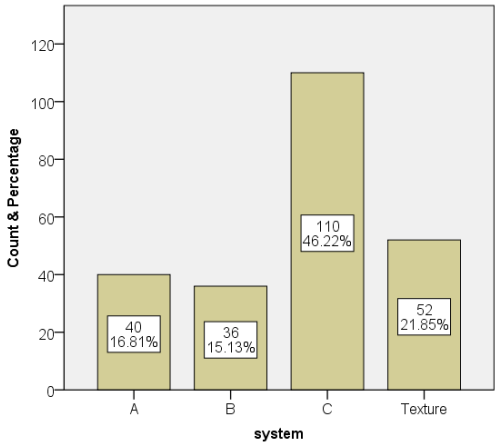


FIGURE 6.21: Frequency of each of four systems being chosen under repeated system selection condition.

that more queries were drawn using **System C** compared to using the other three systems and the frequency of the other three systems being selected differed non-significantly from each other.

It is also found that **male** participants were more likely to use System B than females (Wald $\chi^2(1)$=4.34, p=.037), and more queries for **portrait** painting were drawn using System B than those for landscape painting (Wald $\chi^2(1)$=5.28, p=.022). Surprisingly, however, the GEE model suggests that neither task type nor drawing skills significantly influenced the system choices.

Seasonal Landscape Painting Queries 26 participants completed the 100 queries for seasonal landscape painting search, with each individual completing four seasonal scene drawings except two participants only produced two seasons. As shown in Figure 6.22, System Texture was the most popular system to use for **Autumn** scenery painting search, with 48% of Autumn queries were produced by using System Texture. System A was most frequently used to draw the query for both Spring (34.62%) and Winter (36%) landscape painting.

A GEE model indicates that there were non-significant differences in the frequency of each system being selected among four systems (Wald $\chi^2(3)$=5.28,

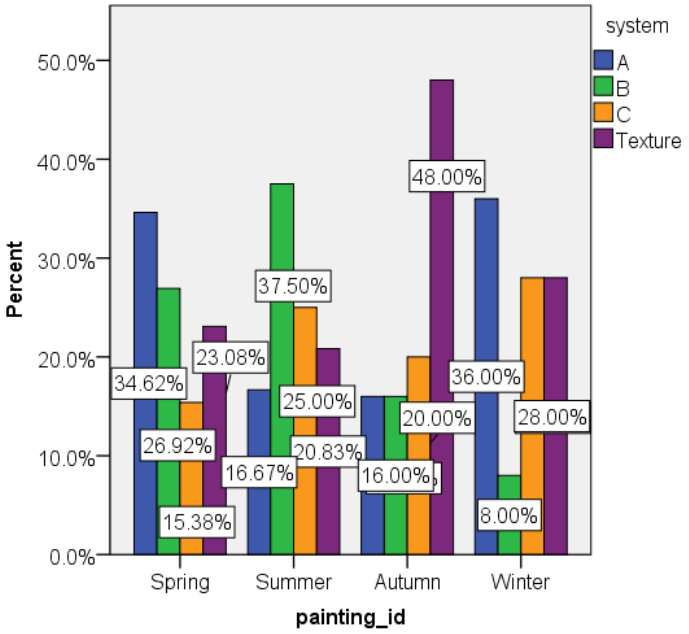


FIGURE 6.22: Frequency of each seasonal scene query made by using four systems under single system selection condition.

p=.153), because each of four systems were equally selected as required by the experiment. The season factor non-significantly affected the system choices (Wald $\chi^2(3)$=2.48, p=.478).

Participants' System Choosing Criteria When making a choice of system for certain tasks, the rule followed by participants is whether one system offers enough choices of colours or textures for representing the expected objects. As Participant 35 stated: "*It (System B) has more types of colours to reflect the colour of spring (so he used System B for the T2-Spring task)... I see Autumn as a season that has many textures which texture brushes might be suitable for; (System) A has simplest range of colours and Winter is simple in terms of colours (so she used System A for the T2-Winter task).*" Participant 65: "*(System) Texture would be good for Autumn's foliage ...(for T2-Autumn).*" Participants avoided choosing System Texture because they would like to use the exact colours mixed by a colour picker.

Some other reasons for choosing one system to use, such as personal preference of system, are excerpted from transcripts of video-logs during participants' drawing procedure:

Participant 49: "*I am familiar with System A, and it is a rush task (T3 task), so I tried to get the better result when using System A.*"

Participant 71: "*System C has various purple colours (for T2-P3 task which requires the purple coat and purple hat); I used my favourite system (System B) to draw the query for my favourite season (for T2-Spring task).*"

6.6.5.4 Use of System Texture

Participants did not adopt the semantic meaning of the brush textures, but rather mainly considered the colour of each brush texture. The common way in which the participants use texture brushes is looking through all texture palettes and selecting one texture that suits their expected colours. The texture self and semantic category are ignored.

Few participants also firstly filled the entire canvas by drawing with expected textural 'background'. Participant 62 even suggested to add the '*background filling*' function to System Texture. Most participants just drew the query with textural brushes without concerning the background. The brush with the largest size was used most frequently when using both system '*Portrait*' and '*Landscape*'.

As most of the participants did not have sufficient drawing techniques, they represented both the portrait painting query and landscape painting query from the view of front or side, top.

6.6.6 Drawing Pattern for Target-unspecific Tasks

As a complex cognitive process, drawings from participants with different background vary with each other, especially when drawing colour sketches. According to [133], for the human being, a fuzzy set of colours would be selected given a colour word. Xue et al. [151] found that the colours from different people's memory of a colour would be located in a compact area.

When a target painting is unspecific (T2 tasks), participants imagined the colour and spatial layout of a portrait or a scene based on their experience when reading a description of target painting. A variety of drawing styles were used by different participants, such as drawing objects from different perspectives, emphasising self-perceived important features and omitting others, and allocating different proportions for foreground objects. Hence, it is no guaranteed that there would be a match of painting in our database for a user's drawn query. The result of the search algorithm is one 'objective' way to measure the query accuracy (search performance), however, this objective accuracy measure depends on the effectiveness of search techniques. Therefore, 'subjective' accuracy measures rated by human evaluators are adopted as necessary supplementary measures to the search performance.

6.6.6.1 Aspect Ratio

It is noted from participants' drawings that aspect ratios of the drawings tend to follow that of the target painting. In other words, portrait painting is depicted in portrait screen direction while landscape painting is in landscape direction. However, for drawing from the description task (T2 task), it is not the case. Almost all queries for portrait painting search were drawn in portrait screen

orientation; while for landscape painting, this is found that 28% of all 100 seasonal scene drawings were presented in a portrait screen orientation, regardless of which system is used.

The current matching metric directly compares the query in 'portrait' orientation to the images in database. Therefore, considering the aspect ratio in the image matching algorithm might produce better results. It is worthy to test this hypothesis in future work. For example, the QbD App could ask participants to select the aspect ratio of canvas before starting drawing a query.

6.6.6.2 Perspective

In line with the work proposed by Van den Broek [15], participants drew portrait painting from a frontal or side view, and all landscape painting queries were produced from a front view. This empirical evidence is in support of the intuition that humans prefer the front view perspective.

For uncommon or complicated portrait paintings from T2 tasks, some participants drew a side view of the subject matter to make the drawing expressive. For example, '*(the girl) reading a book*' in P3 task description is illustrated with side view as demonstrated in Figure 6.23.



FIGURE 6.23: Examples of query with side view: P3 query from Participant 38, 55, 66.

6.6.6.3 Background

Due to the impact of colour interaction [1], that is, the colour perception of one colour may be affected by its surrounding colour(s), it is worthwhile to examine the background of a query. We anticipate that participants select a background colour at the first step of query formulation, then drawing foreground objects on the coloured canvas to reduce the effect of colour interaction.

Background-First Our main insights are that most of the queries were drawn by starting with drawing a background, with 76.89% of T2 tasks' queries and 67.39% of T3 tasks' queries are started with selecting a background colour or drawing a background (entitled as the 'background-first'), and then drawing the foreground on top of background. Only 2.7% of all 483 queries' background are drawn at the end. For both tasks, portrait painting query is more likely to be produced with background-first than the landscape painting query.

Among the 'background-first' queries, most participants firstly pick a background colour by using the background colour picker. It is observed that 62.55% of queries for T2 tasks were drawn by first selecting a background colour, and 63.91% of T3 tasks' queries were also started with the background colour selection first. Hence, this observation might be in support of our QbD system design of separating the foreground colour picker from the background colour picker.

Mixed Background For painting with one fuzzy background colour in T3 task, participants only selected a colour by using the background colour picker; while for painting with complicated background colour, several participants drew the background with brush colour, regardless of selecting a background colour or not. It is found that 4.37% of all queries for T2 tasks and T3 tasks in Experiment 1 and Experiment 2 were made by combining the filled plain background colour with the drawn background, as shown in Figure 6.24. vyeg



FIGURE 6.24: Examples of mixing drawn background with a background colour on queries for T3 tasks.

Complex Background Colour Vague colour words are tested as the background colour in T2-portrait tasks, such as *greenish-yellow* and *bluish-grey*. For these complicated background colours, participants reported their confusion and picked a different background colour or drew a different background by using the brush colour(s). Moreover, it is surprising to find that few participants produced a background with two colours in-between, see Figure 6.25.



FIGURE 6.25: Examples of the complex background of query for task P1 (by female, Chinese Undergraduate), P2 (by female, Chinese Undergraduate), and P4 (by male, English PhD student).

No *Background* Participants tended to draw a query for landscape painting only by using the brush colour(s), which means that no canvas colour is filled.

Almost half of landscape painting queries (49.49%) for T3 tasks were produced by using brush colours only, such as the examples shown in Figure 6.26.

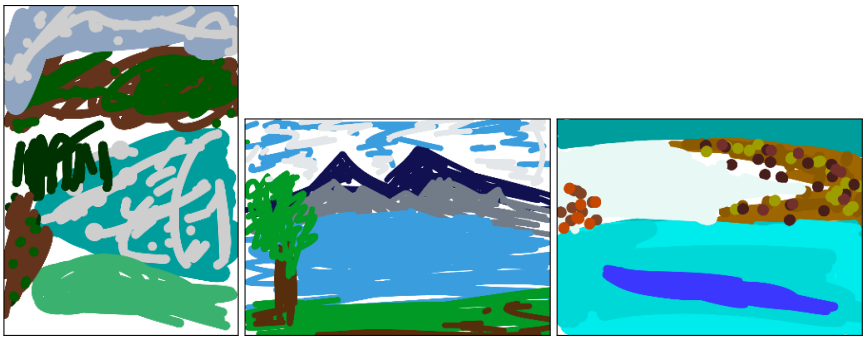
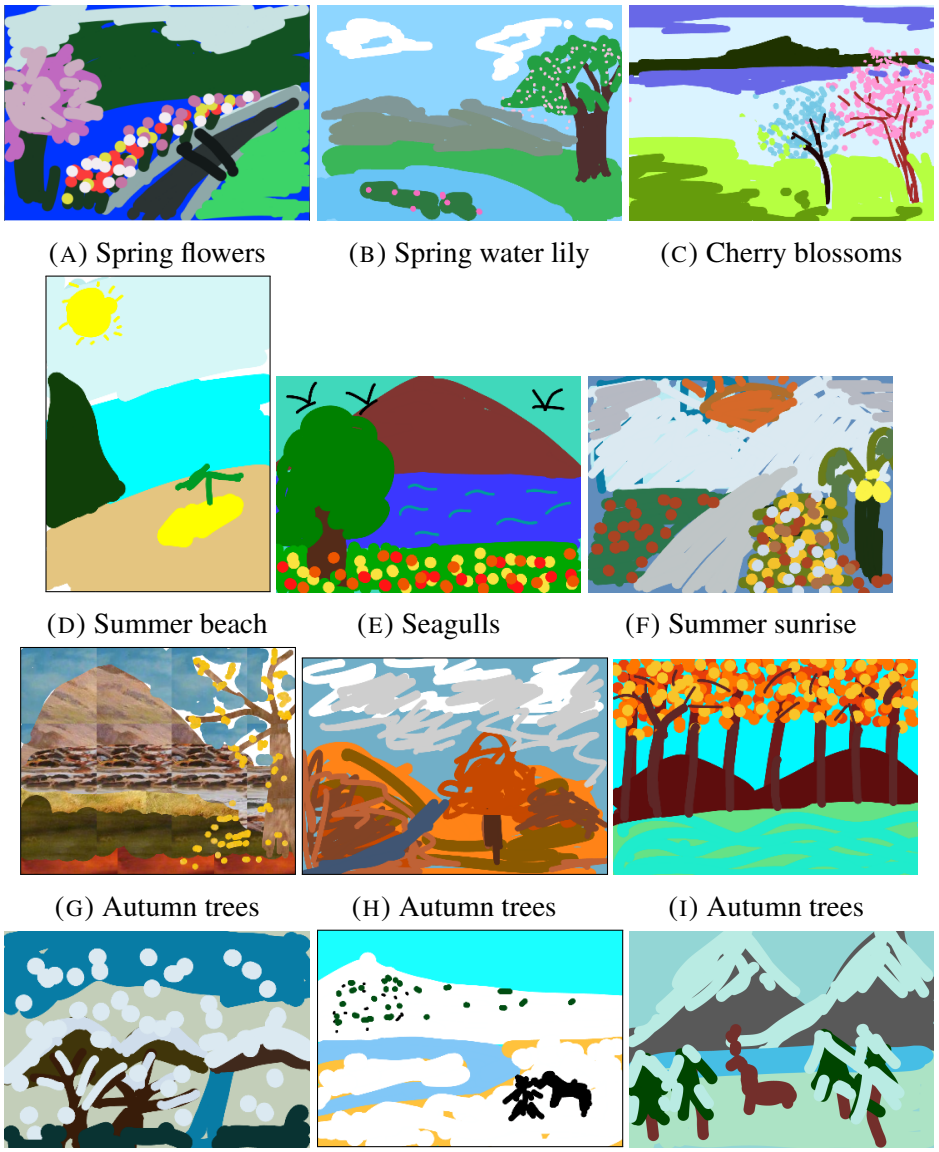


FIGURE 6.26: Examples of landscape queries for T3 tasks made by using brush colours only.

6.6.6.4 Inclusion of Objects

Drawing from Description (T2 tasks) Participants could draw and search for the queries described on the T2 task sheets. For T2-portrait painting search, participants' queries mainly differ in the colours and compositional layout of the objects; while for seasonal landscape painting search, different representative objects were used to depict four seasons by participants with different cultural background and experience. All queries contain at least five natural objects as required - sky, mountain, tree, a body of water, and land.

The informal observations note that participants associated other special objects with a seasonal scene. For example, as shown in Figure 6.27 the flowers and sprouts occur frequently in the query for Spring scene, a beach or greenish theme stands for Summer, yellow leaves and fields represent Autumn, and snow and frozen land or lake are representative for Winter. The inclusions of other natural objects consist of the sun, special trees (e.g., apple tree, coconut palms), specific flowers (e.g., cherry blossoms, water lily), and animals (e.g., seagull, horse, and deer). Artificial objects, such as boat, bridge, on the other hand, were depicted less frequently in the queries.



(A) Spring flowers (B) Spring water lily (C) Cherry blossoms

(D) Summer beach (E) Seagulls (F) Summer sunrise

(G) Autumn trees (H) Autumn trees (I) Autumn trees

(J) Winter snow (K) People & horse (L) Plant & deer

FIGURE 6.27: Examples of the seasonal scene queries, top-bottom rows: spring, summer, autumn, winter.

Drawing after 30-second Viewing (T3 tasks) For tasks of drawing after 30-second viewing of stimuli (T3 tasks), participants could remember and draw highly visual salient objects in the stimuli. However, the details of objects might not be exactly correct, such as colours, size and location of objects in the canvas.

Participants relied on their experience and their understanding, rather than the original colours in the painting. For example, when their memory gets blurred, the associated colours to certain objects would be picked and drawn in a query, e.g., blue sky, skin tone etc. As the number of objects on stimulus paintings increase, participants felt more mental workload to memorise the painting that he or she did not prefer to appreciate.

All participants remembered the primary colour blocks on the stimuli, and small details would be eliminated by human's simplification heuristics [83], producing a sketch geometrically different from the viewed painting. Hence, paintings consisting of simple geometric shapes and topological relationships between elements are easier to remember and draw. Some participants felt that it was easier to remember portrait paintings than landscape ones.

6.6.6.5 Cultural Background and Experience

It is observed that the participants' cultural background indeed influenced their seasonal landscape drawing.

T2-Portrait tasks For portrait painting under target-unspecific tasks (T2 tasks), participants were instructed to produce a drawing for the painting description in English. The fuzzy colour words such as *yellowish-green*, *bluish-green*, *blond* hair, and *brown* skin, etc., are used to encourage participants to explore thoroughly using the colour picker. It is anticipated that participants who reported their first language to be English would draw higher accuracy queries than other

participants. However, the results based on 248 queries under T2 tasks suggest that the participants with different first languages did not differ in query accuracy. But the ethnic group significantly affected the 'subjective' foreground colour accuracy of a query (GEE, Wald $\chi^2(2)$=6.69, p=.035), in that the foreground colours of query from **European** participants was more likely to be rated with higher score than those from Chinese participants (p=.010).

The words used in the task specification might influence the participants' visual imagery which is linked to their colour selection and drawing. The essence behind this effect is different cultural background.

T2-Landscape tasks For example, Chinese participant 51, who is also an expert of East Asian ink wash painting, drew the seasonal landscape queries in East Asian style by using colour pickers as shown by Figure 6.28. He stated that:"*It is enough to put fewest possible lines and shades on the canvas to express its soul*", even he was instructed that the painting database only contains the West oil paintings. He explained during the exit interview that he would like to see the returned results of our QbD App based on his drawings just for fun. The system texture did not allow him to do the East Asian ink wash painting, as shown in Figure 6.28.

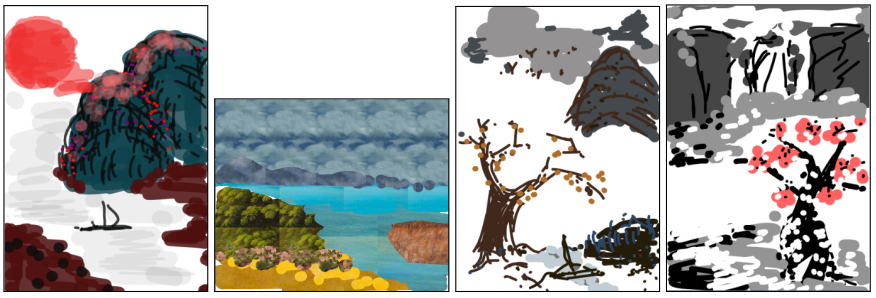


FIGURE 6.28: Queries for four seasonal scenes from Participant 51 (male, Chinese, Engineering PhD student).

By intuitively using their experience, some participants drew the queries for the places he or she had visited or grew up. For example, Participant 67, who is

from India, drew a typical Indian beach scene as shown in Figure 6.29 (left); and Participants 65 drew a beach view of Scotland she had been (see Figure 6.29,right), as she described: "*I am actually thinking of Scottish island, there is beach and sea, but you do get the mountain coming down to the sea on the island. This is the scene saw from the distance... that is the same place I had in the mind.*"

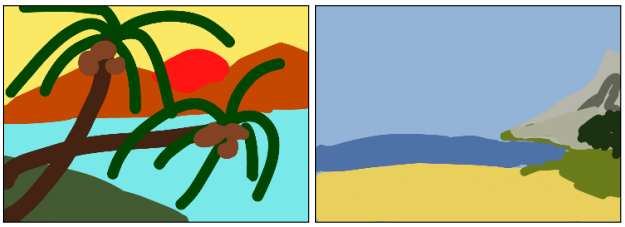


FIGURE 6.29: Summer queries from Participant 67 (female, Indian, Plant & Crop PhD student) and Participant 65 (female, English, HCI research assistant).

Interestingly, some participants produced four seasonal landscape drawing of the same place in the same style, with colour changes indicating different seasons, see Figure 6.30.

T3 tasks - Drawing after 30-second Viewing The T3 task is used to investigate the query drawing performed immediately after seeing the test painting. The ethnic group did not affect all 'subjective' accuracy and search performance of queries.

6.6.7 Participants' Judgements on Relevance

Participants were asked to indicate results relevant to their search intent. Whether an image contains all objects with expected colours in their mind was firstly judged by participants. If no results meet the first requirement, they would choose the images that contain at least one representative object, or the one with similar compositional structure.



FIGURE 6.30: Four season scenery queries with one row from one participant. Top-Bottom: Participant 56 (male, Chinese CS PhD student), Participant 58 (female, Malaysian CS PhD student) and Participant 62 (male, Arabic IT PhD student).

6.6.7.1 Specific Objects or Colours

Participants selecting a result image to be the expected painting mainly depends on whether the image including the special objects or colours in their drawings or in their mind.

For example, as shown in Figure 6.31, Participant 70 selected two images as the relevant results to her query for Summer scene, as she explained: "(In the first matched image) *The sky is fine, and I got the field from the distance.* (In the second match) *Trees are on the right, fields are from the distance and the mountain is over there. But there are more trees than my drawing. These two (images) are acceptable results.*"



FIGURE 6.31: Summer query from Participant 70 and two self-selected most-relevant images.

The **boat** was used as the first relevance criteria by Participant 64, who was satisfied with two images in returned results, as shown in Figure 6.32.



FIGURE 6.32: Autumn query and two self-selected most-relevant images from Participant 64.

The **colour of objects** was also reported as the relevance criterion. For instance, the colours of tree, leaves, and flowers. As seen in Figure 6.33, Participants 70 explained why she chose the image with trees as most matched results: "*This is a kind of idea what I had in my head, but if there was a big mountain there. In terms of the brightness of the colour is closed to my tree leaves.*" This is also happened for Participant 58 when selecting the satisfied result of her Autumn query based on the red leaves.

The **overall colour impression** is also an important factor considered by participants when selecting the relevant results. For example, it is consensus that Winter is about white, and Autumn is about brown or orange. Figure 6.34 and Figure 6.35 list the queries from Participant 72 and Participant 65, who judged the relevance of result image based on the overall colours.

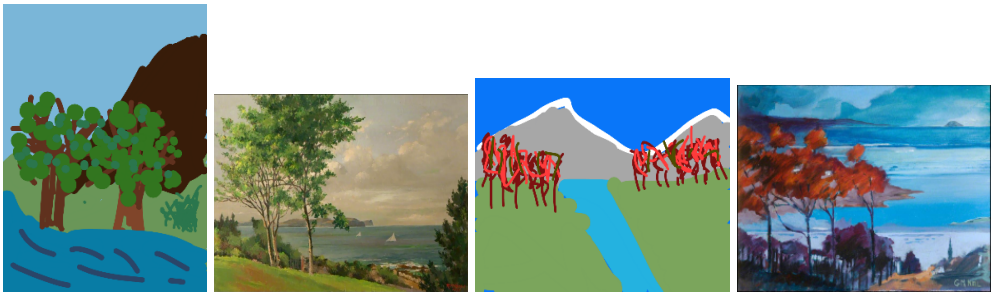


FIGURE 6.33: The '**green tree**' in Spring query and self-selected match from Participant 70 (left two); The Autumn query with '**red leaves**' and self-selected match from Participant 58 (right two).



FIGURE 6.34: A Spring query and self-selected match (left), a Summer query and self-selected match (right) from Participant 72.



FIGURE 6.35: An Autumn query and self-selected match (left), a Winter query and self-selected match (right) from Participant 65.

6.6.7.2 Compositional structure

The compositional structure is another primary criterion for participant's judging the relevance of results. For example, Participant 68 indicated the reason of choosing the relevant image to his query is because '*the river is in between two trees*', see Figure 6.36.

Participant 60 chose the image with '**mountain behind trees**' as the most matched image to his query for both Spring and Winter landscape, see Figure 6.37.



FIGURE 6.36: A Summer query and self-selected match from Participant 68.



FIGURE 6.37: Autumn query and self-selected match (left), Winter query and self-selected match (right) from Participant 60.

6.7 Discussion

The results of Experiment 1 and Experiment 2 suggest that participants were able to draw queries with colour information to search, which supports McDonald et al.'s work of searching on the basis of dominant colours of paintings [93].

6.7.1 Verifying Findings of Experiment 1

The generalisations of findings of Experiment 1 are confirmed with more stimuli and more participants by revealing same effects of system design on the objective measures for copying-tasks. Table 6.10 demonstrates the comparison results of the effects of system design based on a number of different tasks. With the target painting in sight (T1 task), participants spent the **least time** in drawing a query by using the ***discrete-continuous*** **System C**, and **most colours and objects** were depicted in a query when using the ***discrete-discrete*** **System A**, and

the search performance measures on using three systems did not differ significantly from each other. The absence of difference in objective measures between those two experiments suggests that we have successfully selected a stimulus in Experiment 1 to test system effects under the copying-task condition. In other words, with the stimulus in sight (T1 copying-task), one target painting that follows our stimulus-choosing rationale could be enough to evaluate the objective measures.

TABLE 6.10: Comparisons of system effects on the '*Lady*' queries from Ex1 (Effect-Lady), all copying-task queries (Effect-T1) and all tasks on colour systems (Effect-T1∼T3) from Ex1 & Ex2 (n.s.: not significant, '-': not available).

Measures	Effects-Lady	Effects-T1	Effects-T1∼T3
Total time of query formulation	(B, A)> **C**	(B, A) > **C**	**B** > (A, C)
No. of colours depicted in a query	**A** > (C, B)	**A**> (B, C)	(A> C, B)
No. of objects depicted in a query	**A** > (C, B)	**A**> (C, B)	(**A** > B, C)
Number of Top21 query	n.s.	n.s.	n.s.
Number of Top51 query	-	n.s.	n.s.
Colour accuracy of a query	(B, A) > **C**	(**A** > C, B)	(C, A) > **B**
Foreground colour accuracy	n.s.	n.s.	n.s.
Background colour accuracy	n.s.	n.s.	(C, A) > **B**
Least-effort system	(**B** > C, A)	n.s.	-
Most-effort system	n.s.	n.s.	-
Least-time system	n.s.	n.s.	-
Most-time system	n.s.	n.s.	-
Easiest-to-use brush colour picker	(**B** > C, A)	n.s.	-
Hardest-to-use brush colour picker	n.s.	n.s.	-
Easiest-to-use BG colour picker	(B, C) > **A**	(C, B) > **A**	-
Hardest-to-use BG colour picker	n.s.	**A** > (B, C)	-
Most brush colour choices	**B** > **C** > **A**	**B** > (A, C)	-
Fewest brush colour choices	(A, C) > **B**	(A, C) > **B**	-
Most BG colour choices	(C, B) > **A**	(C, B) > **A**	-
Fewest BG colour choices	**A** > (C, B)	**A** > (C, B)	-
Most liked brush colour picker	**B** > (A, C)	**B** > (A, C)	-
Most disliked brush colour picker	n.s.	n.s.	-
Most liked BG colour picker	(B, C) > **A**	(C, B) > **A**	-
Most disliked BG colour picker	(A > B, C)	**A** > (B, C)	-

For copying-tasks, the slight differences of system effects between one portrait stimulus and more stimuli are only found in subjective indicators, including human evaluators' accuracy rankings and participants' usability choices. Firstly, for the colour accuracy of overall query, the '*Lady*' queries made on System C

were rated the lowest accuracy score in Experiment 1; while with more stimuli and more participants, queries for the copying-tasks drawn on System A were assigned higher colour accuracy score than queries made on System C, and this difference between System B and System C was non-significant. Secondly, participants' usability choices of the system with the least-effort consuming and the easiest-to-use brush colour picker was non-significantly different among three systems when considering 72 participants' responses; while 37 participants from Experiment 1 selected System B as the least-effort consuming system and the easiest-to-use brush colour picker more often than System C.

The rest of the subjective measures from more participants are consistent with those of Experiment 1, in that System A's background colour picker was selected to be the hardest-to-use background colour picker and with fewest colour choices. Participants preferred System B's brush colour picker and disliked System A's background colour palette.

TABLE 6.11: Comparisons of colour picker type effects on copying tasks from Ex1 and on all copying tasks from Ex1 & Ex2 (n.s.: not significant).

Measures	Effects-Lady	Effects-4 stimuli
Foreground colour accuracy	n.s.	n.s.
Background colour accuracy	Discrete > Continuous	n.s.
Easiest-to-use brush colour picker	Continuous > Discrete	n.s.
Hardest-to-use brush colour picker	n.s.	n.s.
Easiest-to-use BG colour picker	Continuous > Discrete	Continuous > Discrete
Hardest-to-use BG colour picker	n.s.	Discrete > Continuous
Most brush colour choices	Continuous > Discrete	Continuous > Discrete
Fewest brush colour choices	Discrete > Continuous	Discrete > Continuous
Most BG colour choices	Continuous > Discrete	Continuous > Discrete
Fewest BG colour choices	Discrete > Continuous	Discrete > Continuous
Most liked brush colour picker	Continuous > Discrete	Continuous > Discrete
Most disliked brush colour picker	n.s.	n.s.
Most liked BG colour picker	Continuous > Discrete	Continuous > Discrete
Most disliked BG colour picker	(Discrete > Continuous)	Discrete > Continuous

Table 6.11 shows the comparisons of the colour picker type effects. For the '*Lady*' queries in Experiment 1, the background colour selected by using the discrete colour picker was perceived more accurate than that picked by using

the continuous colour picker. However, it might be arbitrary because System A's background colour palette has the exact background (BG) colour on the '*Lady*'. While considering four stimuli, this difference between the discrete BG colour picker and the continuous BG colour picker is not great.

Except for the choice of the easiest-to-use brush colour picker and the hardest-to-use background colour picker, the effects of colour picker type on participants' usability choices are the same with more participants. 37 participants from Experiment 1 more often chose the continuous colour picker as the easiest-to-use brush colour picker, and their choices of the hardest-to-use background are not different between two types of colour pickers. While with more stimuli, most of 72 participants chose the discrete colour picker as the harder-to-use background colour picker, but the frequency of two colour picker types being selected as the easiest-to-use brush colour picker varied non-significantly.

Overall, participants thought that the continuous background colour picker is easier to use, and they preferred to use the continuous colour picker in both brush and background colour picker groups. The discrete background colour picker is more likely to be selected as most disliked colour picker.

6.7.2 Effects of Task Type

The task type had highly significant effects on both objective and subjective response measures. When a target painting is in sight, more colours and objects were drawn in a query, and consequently it took more time to complete a query for copying tasks than that for other tasks. Queries drawn from the description (T2 tasks) were rated higher foreground colour accuracy and overall colour accuracy than other tasks. Again, the foreground colour accuracy ratings are positively correlated with the entire query's colour accuracy ratings. A query for 30-second present painting (T3 task) was rated the lowest background colour

accuracy score among three types of tasks, even though the details (the number of colours and objects) in a query for T3 task did not much differ from those on a T2 query.

The search performance measures in terms of the number of Top21 queries and the number of Top51 queries were affected by task type, but the effects of task type on search performance are modified by painting type. For portrait painting, queries for T2 tasks get better results than queries for T1 tasks and T3 tasks in terms of the number of Top21 queries and Top51 queries, and more Top51 queries were made for T2 tasks; while for landscape painting, query for T3 tasks obtain worse search performance than that for T1 tasks and that for T2 tasks.

6.7.3 Effects of Painting Type

A query for landscape painting search contains more colours and objects than portrait painting query, but somewhat surprisingly, a portrait painting query took participants more time to complete than landscape painting one.

For the search performance measures, painting type interacts with task type. Only for copying-tasks (T1 tasks), more queries for landscape painting search are successful queries than those for portrait painting, and no painting type differences in search performance are found under target-absent task condition (T2 tasks and T3 tasks).

6.7.4 Effects of Brush Type

The direct effects of brush type are found on all objective and subjective measures. Many more colours and objects were depicted in a query when using the plain colour brush than using the texture brush, but queries drawn on System

Texture were rated more accurate in both foreground colour and query's overall colour impression than those drawn on Colour Systems. Participants produced a similar number of Top21 queries using two types of brushes, but more queries produced using the plain colour brush are Top51-query than using the texture brush. It might because that System Texture provides insufficient colours to make a 'good' query for the search algorithm. Overall, participants indicated that the texture brush is easier-to-use and they preferred to use System Texture when drawing a landscape painting query.

We also found that when using texture brush, participants complete a query quicker than using colour brush. From our observation, participants used a large-size brush more frequently than a small-size brush, which reduces the drawing time. As shown in Figure 6.38, queries made by using System Texture are completed roughly rather than presenting details such as outline, shadow when using the colour brush. That is the way we expect the user do when drawing a query by using our Query-by-Drawing application. Because in real-life, people might not be patient to draw every details in a query, and the image search algorithm adopted by our QbD application is based on the globe-image feature. In other word, to some extend, the texture brush helps the user to ignore the details in a drawn query.

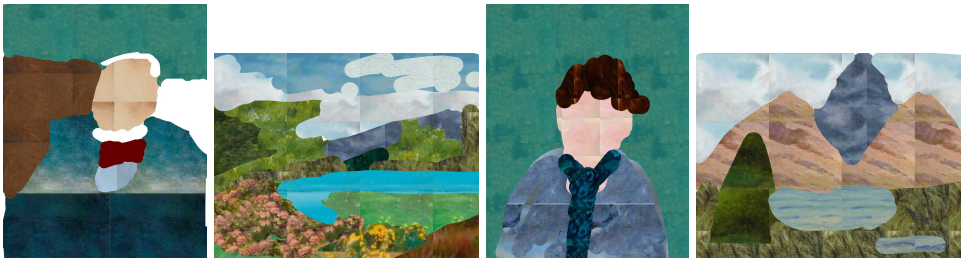


FIGURE 6.38: Examples of queries roughly drawn by using texture brush.

6.7.5 Participant Characteristics

Participant characteristics including gender, ethnic group (Chinese, European, Others) and drawing skill were examined. Because of the attributes of stimuli and task scenarios, we are more interested in the effects of ethnic group on target-unspecific tasks (T2 tasks) and tasks of search for previously-seen paintings (T3 tasks). European participants drew more details such as colours or objects in a query than non-European participants. For target 30-second present tasks (T3 tasks), European participants' queries were rated more accurate than queries of participants from Others ethnic group (non-European and non-Chinese); when drawing a query from the description, European participants produced a query with more accurate foreground than Chinese participants. However, the statistical analyses suggest that the search performance measures were non-significantly affected by participants' ethnic group, regardless of task types.

Results of Experiment 1 and Experiment 2 reveal that **females** drew more details in a query in terms of the number of colours and objects than males, hence females spent longer time in drawing a query than males. Although female's queries were assigned higher accuracy scores than males' by human evaluators, the search performance difference between females and males was non-significant with respect to both the number of Top21 queries and the number of Top51 queries.

Surprisingly, participants' drawing skills did not contribute to search performance. But from a human evaluators' point of view, participants with higher drawing skill produced more accurate queries than those with lower drawing skill. This finding is in line with Ostrofsky et al. [103] who suggested that artists are better at including objects to convey information. Objective measures of query accuracy disagree with subjective measures of accuracy, the reasons might lie in the precision of the matching algorithm or bias from human ratings

of accuracy. Although we observed that few participants with drawing expertise produced complex drawings, as suggested by Case and Okamoto [19], the statistical results in our experiments show non-significant differences in the number of colours and the number of objects in a query from participants with different drawing skills.

Given a freedom to choose any of three colour systems and two texture systems to draw a query for T2-Portrait tasks and all T3 tasks, System C was used most frequently by participants. The statistical analysis reveals that female participants more frequently used System B, which might because females tried to produce more realistic-looking drawing and System B offers full-colour choices for both brush colours and background colours. More queries for portrait paintings were drawn on System B comparing with those for landscape paintings. While for T2-Landscape tasks (search for four seasonal scenes), although we observed that some participants preferred to use System Texture for generating the query for Autumn scene, the statistical results show that the task type and the season had non-significant effects on the system choices.

6.7.6 Drawing Pattern

Most of the queries were drawn by starting with a background first, and the first interaction with our Query-by-Drawing system is selecting a background colour to fill out a canvas or selecting a foreground colour to draw the background. Objects are drawn from the frontal or side view, and almost all portrait painting queries are drawn in 'portrait' direction of phone screen and the majority of landscape painting queries are in 'landscape' screen direction. Still one-third of queries for landscape painting search were drawn in 'portrait' direction, which is the direction fixed by our Query-by-Drawing application.

With the description of target-unspecific portrait paintings, all participants were able to formulate a mental image from a painting's description and draw a simple figure with self-perceived colours of objects. However, different colours were selected by different participants for the same colour words, especially the complex background colour such as *bluish-green*. The cultural background affected participants' drawings of seasonal scenes with five common natural objects sky, mountain, tree, a body of water, and land. For example, an Indian participant produced a tropical beach view as a query for Summer; an English participant drew the Irish island view which is a place she has visited before; and a Chinese participant drew queries for all seasons in style of the East Asian ink wash paintings. With different culture and experience, additional specific objects are also included in the query for seasonal landscape painting. Participants judged the relevance of a result image to their expected target paintings based on the compositional structure of objects and the inclusion of special objects or colours in their drawings or in their minds.

Regarding the target 30-second present tasks, participants could remember and draw highly visual salient objects, while small details of objects are usually eliminated. In case of having vague memory or forgetting, participants rely on their experience and their understanding to draw, rather than the original colours in the painting.

Overall, participants are able to represent their search ideas by drawing colour sketches for both target-specific tasks (copying-tasks, drawing after 30-second viewing) and target-unspecific tasks (drawing from the description of a portrait, drawing seasonal landscape with five objects). Most participants indicated that they were satisfied with the search results of our QbD mobile App, and participants perceived that they completed the tasks successfully. In general, participants were satisfied with the system user interface design, especially the feature of separating background colour picker from the brush colour picker. As Participant 66 put: "*I like the feature that background is not affected by eraser.*"

6.8 Summary

In this chapter, based on a larger sample size (72 participants) and more stimuli (26 paintings) from Experiment 1 and Experiment 2, the effects of system design, task type, painting type, brush type and user characteristics such as gender, ethnic group and drawing skill are investigated on both objective and subjective measures. Participants' query formulation patterns and results judgement criteria for target-unspecific tasks are also explored. From these analyses, a set of basic guidelines about making effective judgements on selecting colour picker and brush type can be derived.

The findings of the effects of system design and colour picker type from Experiment 1 are verified by more stimuli and participants in terms of objective measures and user preference. However, participants' self-reported Likert responses on the usability of colour pickers vary slightly as sample size increased, e.g., participants' attitudes on the system with the least-effort consuming and the easier-to-use brush colour picker.

Task type affects both objective and subjective measures. When producing a query with the target painting in sight (T1 tasks), more details are included in a query and consequently more time is spent on the query formulation than the other two types of tasks. For the target-absent tasks (T2 tasks and T3 tasks), most participants could draw a query from their mental images and they were satisfied with the search results. The query drawn from the description is rated most accurate by human evaluators, and queries for 30-second present painting are perceived to have the least accurate background colour. Given freedom of choosing which system to use, System C is used more often, and whether a system provides the expected colours or textures and the complexity of selecting wanted colour are firstly considered by participants when making a system choice.

The brush type has effects on the time of query formulation and query completeness, in that more colours and objects are depicted in a query and thus more time is spent in drawing a query by the use of plain colour brush. More queries drawn on colour systems are Top51-queries than those drawn on System Texture, while opposite results are obtained from evaluators' foreground colour and overall colour accuracy ratings. The texture brush is perceived easier-to-use than the plain colour brush and participants preferred to use the texture brushes for drawing a query for landscape painting search. Participants did not use the semantics of textures rather an overall colour impression of textures during query formulation.

Most of the queries are drawn in the same orientation of the target painting, e.g., portrait screen orientation is used for drawing a portrait painting query. Participants usually start drawing a query with the background. Some special objects are included in the query for target-unspecific tasks, which depended on the participant's cultural background and experience. These objects or colours or compositional structure depicted in a query or in the head are mainly examined by participants when judging the relevance of search results.

Chapter 7

Experiment 3 - Visual Design Attributes of Continuous Colour Picker

7.1 Introduction

The results of Experiment 1 and Experiment 2 suggest that participants preferred the continuous colour picker in comparison with the discrete colour picker. However, the results also raise some questionable concerns about two continuous background colour pickers with different visual designs: colour ring and colour square. And objective measures of the two continuous colour pickers are not collected. Moreover, the colour picker preference data reveal inconsistent patterns, e.g., some participants thought the colour ring is easier to use than the colour square, and vice versa. In other words, the first two experiments have not shed light on which type of continuous colour picker should be built into a Query-by-Drawing (QbD) mobile application, neither which shape of the input area is helpful, nor which input technique is preferred.

In Experiment 3, five 1D-2D continuous colour pickers are explored with an aim to understand the effects of three visual design attributes on the system usability and task performance: the combinations of three colour parameters, the shape of 1D input area (*bar* vs. *half-ring*) and the shape of 2D input area (*square* vs. *wheel*). The effects of these visual design factors on query formulation and search performance are discussed, as well as participants' self-reported preferences and usability ranking. The preliminary results are reported in [156].

This chapter is structured as follows: in Section 7.2 we firstly list several purposes of Experiment 3. We introduce the database used in Experiment 3 in Section 7.3. In Section 7.4 we present the interface designs of five colour pickers. We describe the experimental design and procedure in Section 7.5. Statistical analyses of three visual design attributes are shown in Section 7.6 and finally some discussions are given in Section 7.7.

7.2 Aims of Experiment

This experiment is designed to address the following research question (RQ):

RQ4 : How do combinations and visual representations of three colour parameters of a 1D-2D continuous colour picker influence users' task performance, preference and usability of a QbD mobile application?

The specific purposes include:

- To examine the effects of combinations of three parameters of a colour on users' task performance and the usability of the QbD App.
- To examine the effects of the slider shape on users' task performance and the usability of the QbD App.

- To examine the effects of the shape of the 2D input area on users' task performance and the usability of the QbD App.

7.3 Database and Results Displaying

The database used in Experiment 3 contains the same paintings as used in Experiment 2 including 400 portrait paintings and 400 landscape paintings from *BBC your Paintings*. The 51 top matched results will be returned to the user after submitting a query.

7.4 Colour Picker Interfaces

As mentioned in Section 2.2.3.2, the most commonly used continuous colour pickers consist of a '1D-2D' input area, with the shape of a ring or a bar slider representing a 1D input area, and the shape of square or a wheel representing a 2D input area.

For '1D-2D' interfaces, there are three possible combinations of three parameters of a colour in HSV colour model: H-SV, S-HV, and V-HS. The *bar-square* interface is used to represent the '1D-2D' colour picker. In other words, a vertical slider bar displays and manipulates the '1D' input area and a square displays and controls the '2D' input area. Thus three systems or colour pickers of three combinations are named as ***H-SVsquare***, ***S-HVsquare***, and ***V-HSsquare***.

As for the shape of 1D input area, firstly, we implement the representation of 2D input area as a HS (*hue-saturation*) colour wheel, and then we build a *bar* slider and a *half-ring* slider of V (*value*) along with the HS colour wheel. The colour pickers generated are entitled as ***V-HSwheel*** and ***Vring-HSwheel*** colour picker.

For exploring the shape of 2D input area, the 1D area is built into a bar slider and the representation of 2D HS (*hue-saturation*) area will be compared: *square* vs. *wheel*, which is formed into ***V-HSsquare*** vs. ***V-HSwheel*** colour picker.

To keep the ease-of-use of our QbD App and to explicitly compare different designs of continuous colour pickers, the brush colour picker and background colour picker are still separated from each other. But different from the systems used in previous experiments, both brush colour picker and background colour picker adopt the same user interface design.

For each colour picker interface, a set of boxes representing 20 recently used colours are included in chronological order, with the leftmost colour box presenting the most recently applied colour. Two colour swatches are displayed at the bottom, with the colour swatch on the left showing the last used colour, and the colour swatch on the right serves as a temporary holder for the currently selected colour. When users press the right colour swatch, the brush or background colour will be changed to the new colour; when users touch the left colour swatch, both the handle and the indicator in the 2D area would move to the status of last used colour. For all colour pickers, the initial colour value is HSV (255, 0, 0), which is completely black.

7.4.1 H-SVsquare Colour Picker

The *H-SVsquare* colour picker (see Figure 7.1) is an extension of System C's background colour picker, with more display elements: the history swatches, a 2D input area and a vertical *hue* slider. The dynamically coloured 2D square allows users to simultaneously access both *saturation* and *value* (aka *brightness*), while *saturation* related to the X-axis and *value* related to the Y-axis. The *hue* slider is shown statically during colour selection no matter what the current values of *saturation* and *brightness* are. The *hue* values are ordered from

0 to 255, where '0' is the top of the slider (pure red) and '255' was the bottom of the *hue* slider (pure red again). When a *hue* value is selected, the 2D SV (*saturation-value*) square will be changed accordingly (please refer to the 'dynamically coloured' in [145]). Specifically, the SV colour square fades from white to the selected *hue* horizontally and from the selected *hue* to black vertically.

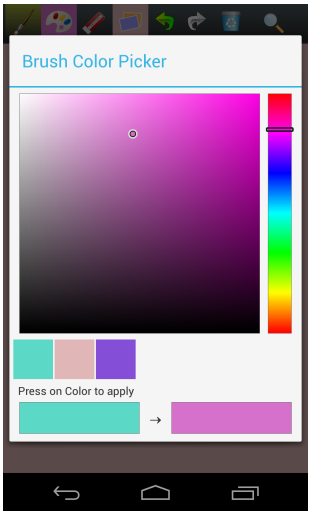


FIGURE 7.1: Screenshot of of ***H-SVsquare*** colour picker interface.

7.4.2 S-HVsquare Colour Picker

The *S-HVsquare* colour picker contains a saturation slider and an HV (*hue-value*) square input area. The HV square consists of the *hue* shown on the X-axis, and the *value* displayed on the Y-axis. Users can select a colour based on its *hue* by dragging the small circle in the HV square area, and then the final colours can be adjusted by dragging the right vertical *saturation* slider. The 2D input area is static while the vertical *saturation* slider is dynamically coloured, from the selected *hue-value* colour to the fully-saturated colour, see Figure 7.2.

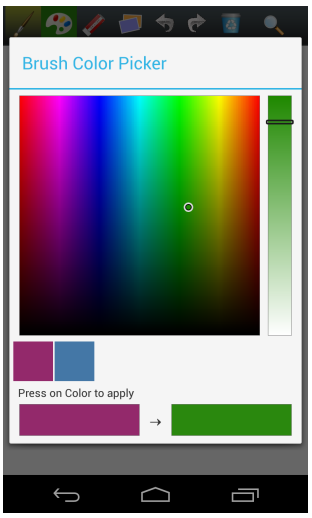


FIGURE 7.2: Screenshot of of ***S-HVsquare*** colour picker interface.

7.4.3 V-HSsquare Colour Picker

The *V-HSsquare* colour picker (see Figure 7.3) comprises a *value* slider and a HS (*hue-saturation*) colour square. Similar to the *S-HVsquare*, the 2D input area in *V-HSsquare* is used to simultaneously control the *hue* and *saturation*, and the vertical *value* slider is dynamically coloured from the selected *hue-saturation* colour to black.

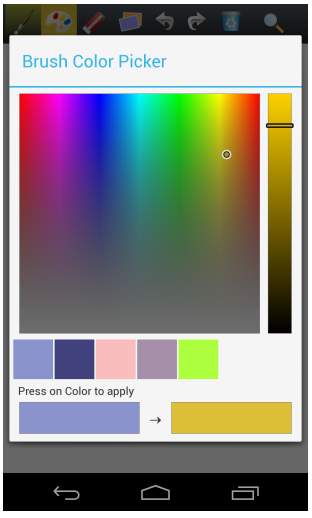


FIGURE 7.3: Screenshot of of ***V-HSsquare*** colour picker interface.

7.4.4 V-HSwheel Colour Picker

As shown in Figure 7.4, the *V-HSwheel* colour picker includes a static HS (*hue-saturation*) colour wheel and a horizontal *value* slider bar. The HS wheel is arranged in polar coordinates with the *hue* varying with the angle and the *saturation* varying with the radius. The *value* slider bar is ranged from black (0) to the selected colour. When the cross on the colour wheel is moved, the right end of the *value* slider changes dynamically with the selected *hue-brightness* colour.

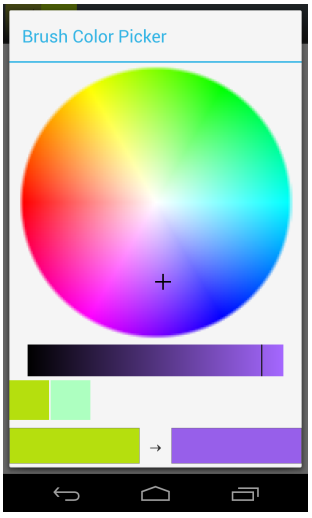


FIGURE 7.4: Screenshot of of ***V-HSwheel*** colour picker interface.

7.4.5 Vring-HSwheel Colour Picker

The coloured mechanism of *Vring-HSwheel* colour picker is similar to *V-HSwheel* colour picker, except that the *value* slider is represented by a *half-ring* on the right and the chosen colour is shown on the left-hand *half-ring*, see Figure 7.5.

Users can pick a *hue* with expected *saturation* from the central HS (*hue-saturation*) colour wheel by dragging the inner small circle around. The chosen colour will be shown on the left *half-ring*. The *value* is set by a positioned handler with

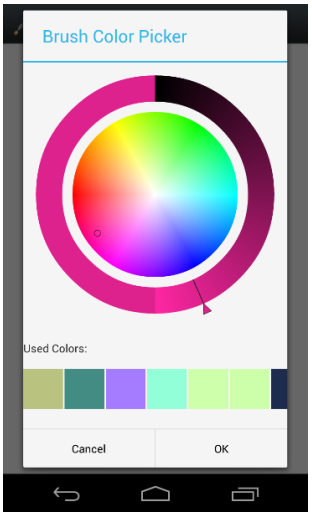


FIGURE 7.5: Screenshot of of ***Vring-HSwheel*** colour picker interface.

a small triangular colour indicator in the area of the right *half-ring*. The *value* ranges from the current selected *hue-saturation* colour to black. The central colour wheel is static and the half-annular slider on the right dynamically shows the shades of the current *hue* and *saturation*.

7.5 Experimental Design and Procedure

From the lessons learnt from previous two experiments, we forced each participant to draw a query within a limited amount of time, so that participants have to concentrate on the key features in the painting. Apart from the different user interfaces, Experiment 3 has the same design as Experiment 2.

7.5.1 Stimuli

The stimuli are selected from Experiment 2, as shown in Figure 7.6 including the '*Sitman*', the '*Flowers*', four T2-portrait stimuli except the '*P3*', and four seasonal landscape paintings. The '*P3*' is excluded from stimuli because almost

all participants from Experiment 2 indicated it is too difficult to draw a query for '*P3*' and only 37.5% of '*P3*' queries are successful.



(A) *Man* (B) *Flowers* (C) *P1* (D) *P2* (E) *P4*

FIGURE 7.6: The stimuli with names used in Experiment 3. (A)-(B) are used for T1 tasks, (C)-(E) are the target painting for T2-portrait tasks.

7.5.2 Participants

Forty-one college students (38 female, mean age 19 years, age range=17-29) have volunteered to take part in the experiment. All participants reported having experience of using touchscreen mobile phones and none of them self-indicated themselves as drawing-experts or artists. All participants are novice users of our QbD Apps and right-handed.

More than a half (32) of the participants are taken from a pool of students enrolled in a nursing course, while the remainder are recruited from students enrolled in a medical imaging course (5) and a dental course (4). All students received a thanks-gift for completing the experiment and their participations are entirely voluntary in all stages of the experiment.

7.5.3 Procedure

The instructions and training phase of Experiment 3 are identical to those of Experiment 2. Participants were asked to fill out the pre-questionnaire (see Appendix A) and the same practice session was conducted to allow participants

getting familiar with the five systems before doing the formal tasks.

All participants were asked to draw a query by copying one portrait painting and one landscape painting on each of five systems, and to rate in the questionnaire **Q2** (see Section 4.5.3.2) when completing one drawing on one system before submitting their queries for search. This rule reduces the impact of search results from matching algorithm on participants' perception of testing system. We asked participants to indicate the best worst system among each of three interface groups: three *bar-square* interfaces, two wheel-based interfaces, and group of bar-square and bar-wheel interfaces.

For time-unlimited T2 tasks and T3 Tasks, firstly, participants were allowed to use any system for conducting any one T2 task or T3 task. Secondly, the number of queries for T2 tasks and T3 tasks are reduced due to the practical restriction, e.g., limited time and resources. Participants were randomly allocated one T2-portrait painting search, and two portrait and two landscape paintings for T3 tasks, and they were requested to search for one seasonal landscape painting (T2-landscape task) by drawing a query with five objects at least - sky, mountain, river, tree and land - on their favourite system. After completing each query for T2 tasks or T3 tasks, participants were required to rate in the questionnaire **Q4** (see Section 4.5.3.2).

7.6 Results

Statistical methods are adopted for the factor effects analyses, and the exploratory variable that shows a non-significant effect will not be reported. Since the number of attempts a participant refining his or her query cannot be controlled, some participants refined their initial queries several times, while others conducted no refinement at all. Hence, only the initial queries for each task are considered in the result analyses.

The outcome variables comprise the total time of query formulation, the quality of the query (e.g., the number of colours and objects depicted on a query, the query accuracy ratings), and the search performance (the number of Top51 queries). Since the same colour picker interface is used for implementing the brush colour picker and the background colour picker, only the overall colour accuracy of each query is rated by human evaluators. Because of the positively-skewed distributions of continuous outcome measures and multiple measures collected from each participant, the *Generalised Estimating Equations* (GEE) models are used to explore the effects of the colour parameter combinations, the shapes of 1D and 2D input area, and participant characteristics on query formulation and system usability.

The Friedman test or Wilcoxon signed-rank test is adopted to analyse the difference in participant's usability ratings among colour picker designs at a significance level of α=0.05. If the Friedman test shows a statistical significance ($p<.05$), a post-hoc analysis with pairwise Wilcoxon signed-rank tests will be conducted with a Bonferroni correction applied, resulting in a significance level at α=0.017.

7.6.1 Colour Parameter Combinations - H-SV, V-HS, S-HV

7.6.1.1 Time of drawing a Query

177 data points are available for analysing the time for drawing a query by using a GEE model. The results reveal that colour picker design (three colour parameter combinations) had non-significant effects on the time spent in forming a query (Wald $\chi^2(2)=1.25$, $p=.534$).

There was a highly significant effect of task type on the query drawing time (Wald $\chi^2(2)=25.12$, $p<.0001$), with much more time spent on a query for T1

task (M=338.22s, SE=15.98) than for T2 task (M=222.05s, SE=12.89) ($p<.0001$) and T3 task (M=242.37s, SE=21.67) ($p<.0001$). This might be because participants attempted to draw more details in the painting if the target painting is continuously presented. The time difference between queries for T2 task and those for T3 task was non-significant ($p=.465$), as shown in Figure 7.7 (left).

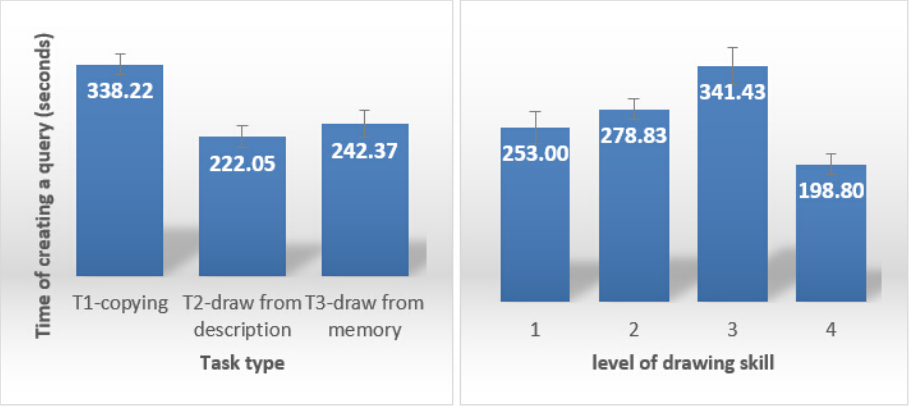


FIGURE 7.7: Main effects of task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean.

The drawing skill had a highly significant effect on the time of query formulation (Wald $\chi^2(3)=30.02$, $p<.0001$), in that participants with higher self-rated drawing skill (score of 4) spent shorter times (M=198.80s, SE=11.78) drawing a query than those with lower level of drawing skill ($p<.05$), and participants with moderate drawing skills (score of 3) spent the longest time ($p<.01$), as shown in Figure 7.7 (right).

7.6.1.2 Number of Colours Depicted in a Query

A GEE modelling 183 data points reveals that the number of colours depicted in a query was affected significantly by the colour picker design (Wald $\chi^2(2)=9.33$, $p=.009$). Pairwise comparisons indicate that fewer colours were used to draw a query by using system *S-HVsquare* (M=6.83, SE=0.24) compared with that using system *H-SVsquare* (M=7.48, SE=0.27) ($p=.006$) and using system *V-HSsquare* (M=7.38, SE=0.28) ($p=.015$). The difference in the number of colours

in a query made by using *H-SVsquare* and *V-HSsquare* was non-significant (p=.653), as shown in Figure 7.8 (left).

Task type had a significant effect on the number of colours in a query (Wald $\chi^2(2)$=14.71, p=.001), in that participants drew more colours in a query for T1 tasks (M=8.29, SE=0.38) than that for T2 tasks (M=7.05, SE=0.36) (p=.010) and for T3 tasks (M=6.45, SE=0.36) ($p<.0001$). This difference between T2 tasks and T3 tasks was non-significant (p=.241), see Figure 7.8 (middle).

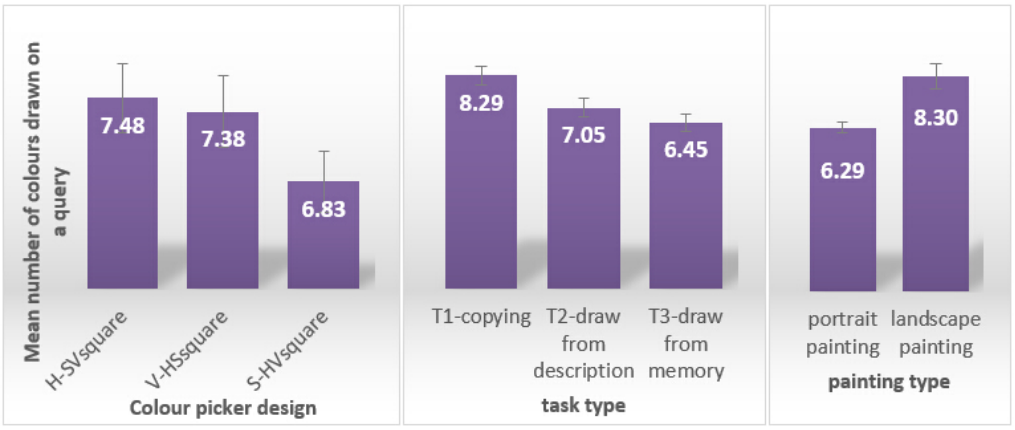


FIGURE 7.8: Main effects of task type and drawing skill on the number of colours depicted in a query. Error bars are the standard error of the mean.

Painting type also highly significantly affected the number of colours ($\chi^2(1)$=13.59, $p<.0001$), with more colours were depicted in a landscape painting query (M=8.30, SE=0.50) than in a query for portrait painting search (M=6.29, SE=0.21).

7.6.1.3 Number of Objects Depicted in a Query

The GEE model analysis indicates that there was a non-significant main effect of colour picker design (Wald $\chi^2(2)$=0.07, p=.964) on the number of objects drawn in a query. The task type, painting type, and participant's gender and drawing skill did not affect the number of objects depicted in a query.

7.6.1.4 Search Performance

There are 183 queries submitted to our Server. The query getting the target painting in top 51 most matched results is defined as the Top51-query. The dependent variable *pos_51* is coded as '1' if a query is a Top51-query, otherwise coded as '0', indicating that the query hits the target painting out of top 51 matched results.

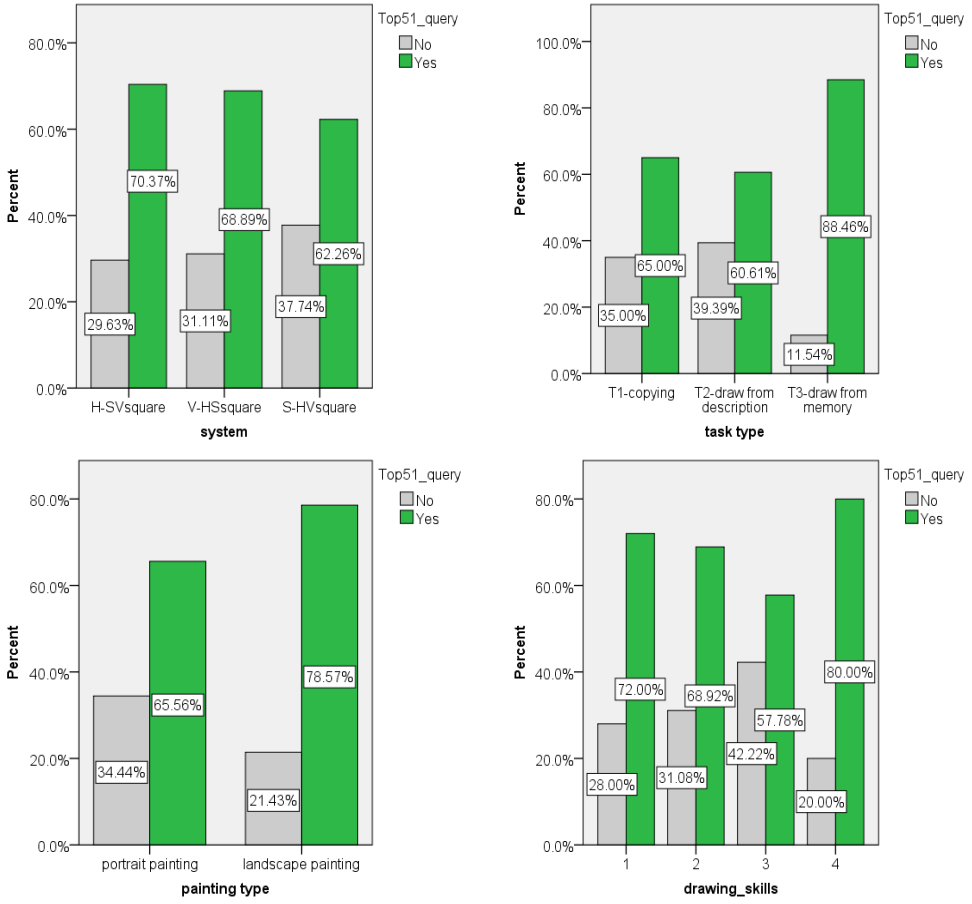


FIGURE 7.9: Percentages of the Top51 query and non-Top51 query with respect to colour picker (systems) design, task type, painting type, and drawing skill.

Overall, 66.12% of queries drawn by using the three colour pickers successfully hit the target painting. As can be seen in Figure 7.9, 70.37% of queries drawn on system ***H-SVsquare*** are Top51 queries, followed by 68.89% of queries made on system *V-HSsquare* and 62.26% of queries produced on *S-HVsquare*.

Most queries for T3 tasks (88.46%) are successful, and 65.56% portrait painting queries and 78.57% of landscape painting queries successfully return the target paintings, while 80% of queries from participants with the highest drawing skill (score of 4) are successful.

However, the GEE model reveals that the colour picker design had non-significant effects on the number of Top51 queries (Wald $\chi^2(2)=1.01$, $p=.602$). Other factors, including task type, painting type and drawing skill, non-significantly influence the number of Top51 queries.

7.6.1.5 Human Ratings on Colour Accuracy

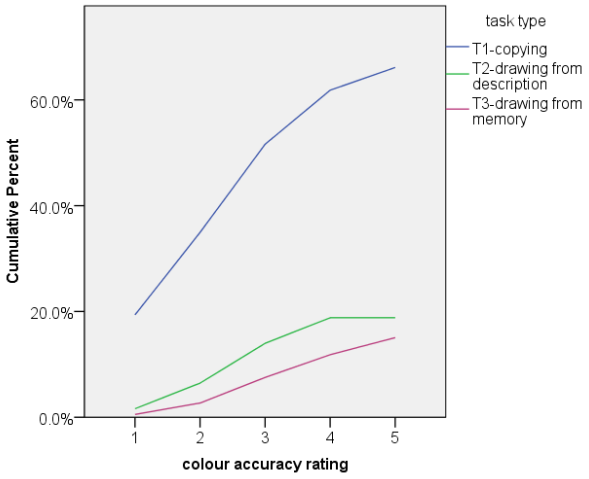


FIGURE 7.10: The observed cumulative colour accuracy ratings of queries made by using three *bar-square* colour pickers.

186 queries are produced using three systems. A GEE with multinomial distribution suggests that the colour picker design was non-significantly related to the colour accuracy ratings of a query (Wald $\chi^2(2)=0.60$, $p=.739$).

Task type was significantly associated with the colour accuracy ratings (Wald $\chi^2(1)=11.31$, $p=.004$). Queries for T2 tasks were less likely to be rated higher than queries for T3 tasks ($p=.018$), and queries for T1 tasks were less likely to

be rated higher than those for T3 tasks (p=.001). The difference in colour accuracy ratings of queries for T1 tasks and T2 tasks was non-significantly different (p=.196), as shown in Figure 7.10.

7.6.1.6 Participants' Usability Ratings

The ranking data on participants' perceived usability is collected from questionnaire **Q2** (see Section 4.5.3.2), including the *relaxation of use, ease of use, complete speed, ease-of-picking a colour, enough colour choices*, and *quick-to-select a colour*. The self-rated similarity of drawing to the original painting is also included. Each participant was requested to respond on a 7-point scale from worst to best for each of these usability features after completing a T1 task on each system.

Friedman tests are conducted to explore whether there are statistically significant differences in these usability features between the three colour picker tools. The post-hoc analyses using Wilcoxon tests (at a significance level α=0.017) will be conducted if the results from Friedman test are significant ($p<.05$).

TABLE 7.1: Results of Friedman tests, Median and IQR of 7-point ratings of usability of three HSV-combination colour pickers.

Usability features	Significance of Friedman test	Median (IQR)		
		H-SV	V-HS	S-HV
Relaxation	$\chi^2(2)$=2.05, p=.359	4 (3-6)	4 (3-5)	5 (2.5-5)
Ease-of-use	$\chi^2(2)$=0.59, p=.746	4 (3-6)	4 (3-5)	5 (3-6)
Quick to complete a query	$\chi^2(2)$=2.90, p=.865	4 (3-5.5)	4 (3-5)	4 (3-5.5)
Easy to get wanted colour	$\chi^2(2)$=2.95, p=.229	4 (3-6)	4 (3-5)	4 (3-5)
Quick to pick a colour	$\chi^2(2)$=0.94, p=.625	4 (3-5)	4 (2.5-5)	4 (3-5)
Enough colour choices	$\chi^2(2)$=4.91, p=.086	6 (5-7)	5 (4-6)	5 (4-6)
Similarity of query	$\chi^2(2)$=1.36, p=.508	3 (2-4.5)	4 (2.5-5)	4 (2-5)

As demonstrated in Table 7.1, there were non-significant differences ($p>0.05$) in usability features between the three colour pickers. Participants perceived

that their drawings were approximately similar to the target painting regardless of which colour picker is used.

7.6.1.7 Preference and Best-Worst Choices of Colour Picker

The results of a Friedman test suggest that there were no statistically significant preference differences among three colour pickers ($\chi^2(2)$=5.29, p=.071), see Figure 7.11.

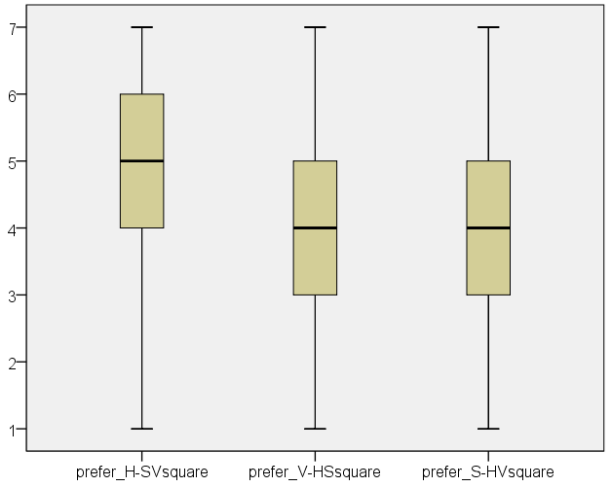


FIGURE 7.11: The box-plot of participants' 7-point preference ranking scales for three *bar-square* colour pickers.

123 best-worst choices of *bar-square* colour pickers were collected from 41 participants. As shown in Figure 7.12, 56.10% of participants chose the *H-SVsquare* as the '*best-to-use*' colour picker and 39.02% of the '*worst-to-use*' colour picker responses are related to the *S-HVsquare* colour picker.

A GEE model analysis of the '*best-to-use*' colour picker data indicates that the participants' choices of the '*best-to-use*' colour picker were significantly different among three colour pickers (Wald$\chi^2(2)$=7.94, p=.019). The **H-SVsquare** colour picker was more frequently to be chosen as the '*best-to-use*' colour picker than the *V-HSsquare* (p=.003) and the *S-HVsquare* (p=.016), and the difference

in choosing between the *V-HSsquare* and the *S-HVsqaure* was not significant (p=.637). The results of another GEE model on the '*worst-to-use*' colour picker choices reveal that there was a non-significant difference in the '*worst-to-use*' colour picker among three *bar-square* colour pickers with HSV combinations (Wald $\chi^2(2)$=0.61, p=.737).

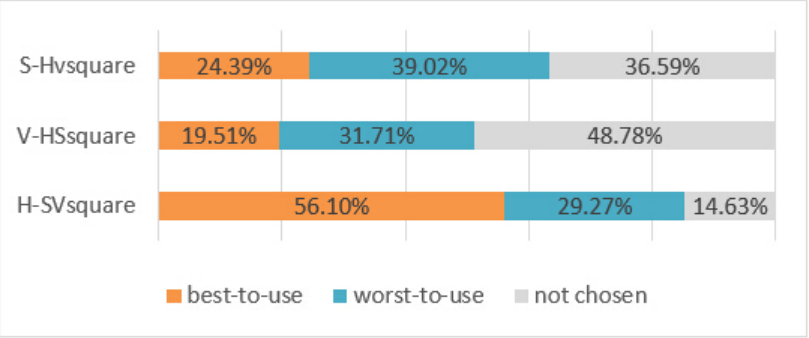


FIGURE 7.12: Proportion of colour picker selected as the '*best-to-use*' and '*worst-to-use*' *bar-square* colour picker.

7.6.2 Shape of 1D Input Area - Bar vs half-Ring

7.6.2.1 Time of drawing a Query

122 data points are available for analysing the time of drawing a query. A GEE model reveals that the colour picker design non-significantly affected the time of drawing a query (Wald $\chi^2(1)$=2.06, p=.151).

Painting type significantly influenced the time of query formulation (Wald $\chi^2(1)$=7.95, p=.005), with participants spending less time on drawing a query for portrait painting search(M=223.38s, SE=14.70) than for landscape painting search (M=217.58s, SE=28.73). The task type also significantly affects the query formulation time (Wald $\chi^2(2)$=8.85, p=.012), in that the query formulation time was much longer for conducting a T1 task (M=330.53s, SE=29.08) that for a T2 (M=229.58s, SE=19.53) and T3 task (M=249.01s, SE=22.43), as shown in Figure 7.13.

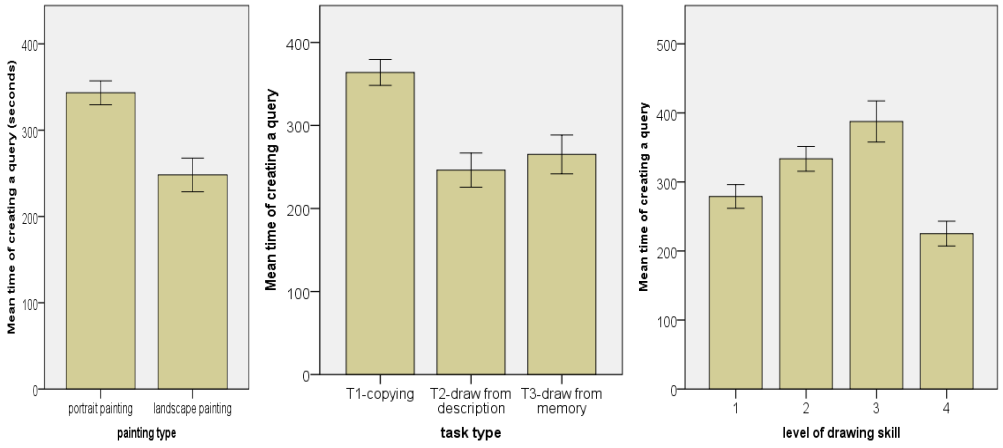


FIGURE 7.13: Main effects of painting type, task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean.

There was a significant effect of drawing skills on the time of query formulation (Wald $\chi^2(3)$=8.27, p=.041), in that participants with the highest self-rated drawing skill (score of 4) spent significantly shorter time (M=217.12s, SE=18.72) than those with drawing skill score of 1 (M=285.49s, SE=20.07) (p=.009) and those with drawing skills scores of 2 (M=280.52s, SE=22.19) (p=.018) and those with drawing skills scores of 3 (M=289.43s, SE=33.74) (p=.059).

7.6.2.2 Number of Colours Depicted in a query

126 data points are analysed by a GEE model, which suggests that the colour picker design had non-significant effects on the number of colours used for drawing a query (Wald $\chi^2(1)$=0.65, p=.419).

There is a highly significant effect of painting type on the number of colours drawn in a query (Wald $\chi^2(1)$=17.53, $p<.0001$), with more colours depicted in landscape painting queries (M=8.69, SE=0.47) than in portrait painting queries (M=6.04, SE=0.24). The task type also affected the number of colours (Wald $\chi^2(2)$=8.44, p=.015). Pairwise comparisons suggest that more colours are used to draw a query for T1 task (M=8.23, SE=0.38) than a query for T2 task (M=7.19, SE=0.38) (p=.015) and a query for T3 task (M=6.42, SE=0.35) (p=.005), and

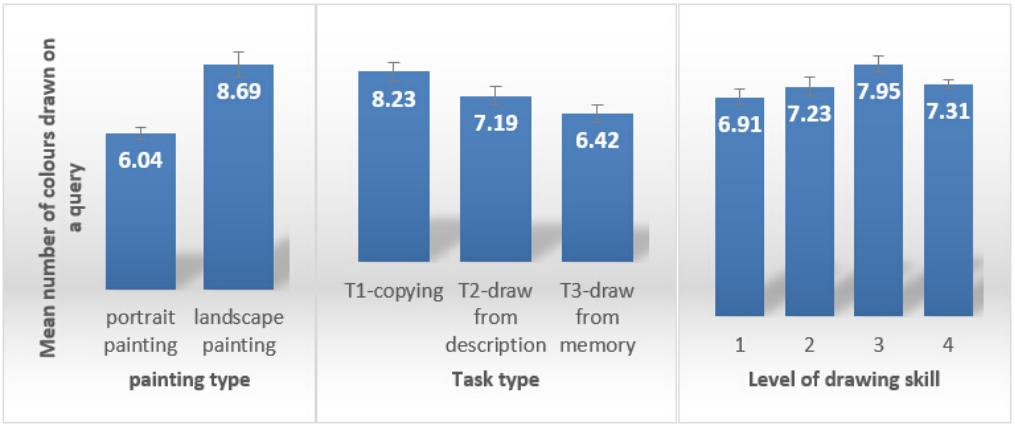


FIGURE 7.14: Main effects of painting type and task type on the number of colours depicted in a query. Error bars are the standard error of the mean.

difference between T2 task and T3 task was not significant (p=.073), as shown in Figure 7.14.

The participants' drawing skill also had a significant effect on the number of colours drawn in a query (Wald $\chi^2(3)$=11.78, p=.008). Participants with drawing skill score of 3 used the most colours for drawing a query.

7.6.2.3 Number of Objects Depicted in a query

A GEE model analysis based on 126 data points reveals that the colour picker design did not affect the number of objects drawn in a query (Wald $\chi^2(1)$=2.13, p=.144).

There was a significant effect of task type on the number of objects in a query (Wald $\chi^2(2)$=20.94, p<.0001), with more objects drawn in a query for T1 task (M=8.85, SE=0.43) than for T2 (M=7.12, SE=0.26) (p=.001) and for T3 tasks (M=7.02, SE=0.52) (p=.028), while the difference between T2 and T3 task was not significant (p=.866), see Figure 7.15.

Painting type had highly significant effects on the number of objects depicted in a query (Wald $\chi^2(1)$=20.11, p<.0001), in that more objects were drawn in

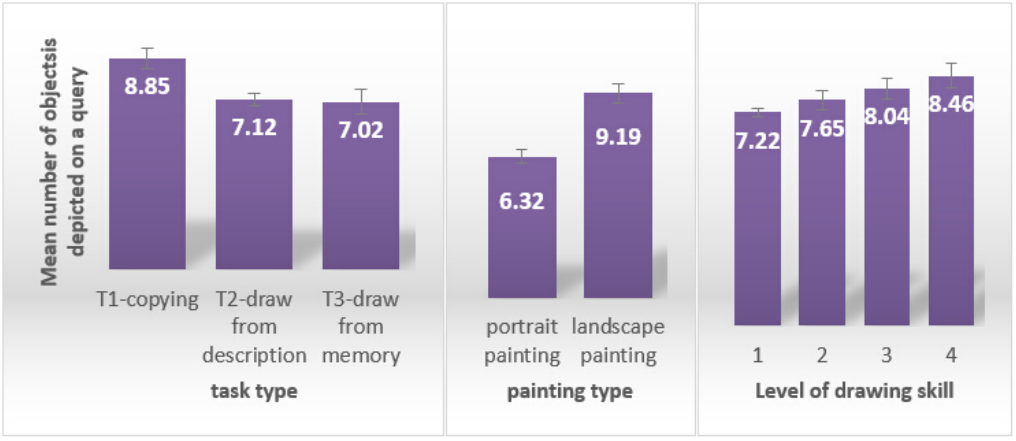


FIGURE 7.15: Main effect of task type, painting type and drawing skill on the number of objects depicted in a query. Error bars are the standard error of the mean.

a landscape painting query (M=9.19, SE=0.43) than that in a portrait painting query (M=6.32, SE=0.32).

Drawing skills were also significantly relevant to the number of objects drawn in a query (Wald $\chi^2(1)$=13.249, p=.004). Participants with higher drawing skill produced more objects in a query than those with lower drawing skill, as shown in Figure 7.16.



(A) *DS=3* (B) *DS=1* (C) *DS=4* (D) *DS=2*

FIGURE 7.16: Drawings created by participant with higher drawing skills (DS, 1=lowest, 5=highest) in Experiment 3.

7.6.2.4 Search Performance

135 drawings were created and submitted to our server for searching. However, only 126 queries hit the results returned because 9 queries crashed due to poor

network connection. The query getting the target painting in 51 most matched results is defined as the Top51-query.

Overall, 67.41% of queries successfully hit the target painting in returned 51 results. As seen from Figure 7.17, 72.15% of queries drawn on the System ***Vring-HSwheel*** are successful, compared to 72.34% of queries made on the System *V-HSwheel* are Top51-query. More Top51 queries are formed for **T3** task (96.30%) than those for T2 task (67.86%) and T1 task (67.79%). More queries for **landscape** paintings (86.67%) are Top51-queries than those for portrait painting search (67.71%). Surprisingly, all queries from participants with the highest level of drawing skills are Top51 queries, followed by 75.56% of queries made by participants with drawing skills score of 1.

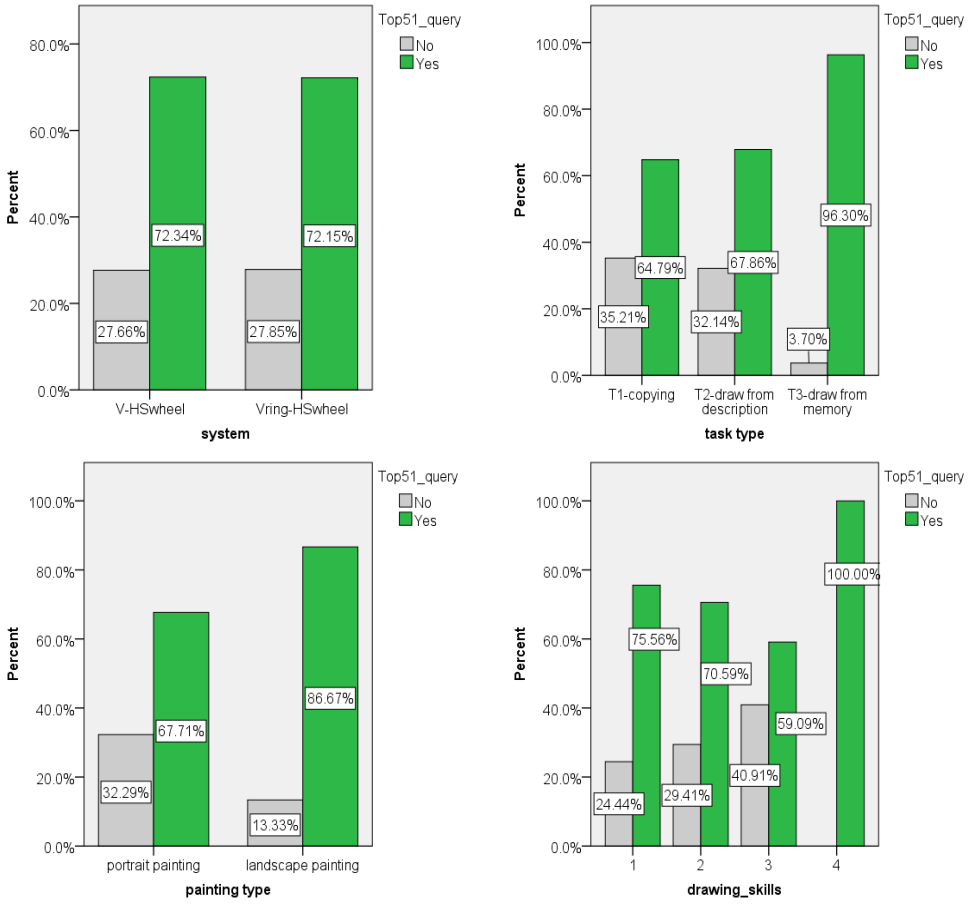


FIGURE 7.17: Percentages of the Top51 query and non-Top51 query in terms of two wheel-based colour pickers, task type, painting type, and drawing skill.

A GEE model analysis reveals that the colour picker design had non-significant effects on the number of Top51 queries (Wald $\chi^2(1)$=0.18, p=.692). In other words, the shape of the slider of colour picker design non-significantly affected the search performance. The painting type and drawing skill did not significantly influence the search performance.

The task type had a significant effect on the search performance (Wald $\chi^2(2)$=6.46, p=.040). However, the pairwise comparisons of the number of Top51 queries among three task types indicate that more successful queries were drawn for T3 tasks compared to those for T1 tasks ($p<.0001$) and for T2 tasks (p=0.004), and the search performance difference between T2 task and T1 task was non-significant (p=.746), regardless of which colour picker is used.

7.6.2.5 Human Ratings of Colour Accuracy

Overall, there are 127 queries drawn using the *V-HSwheel* and *Vring-HSwheel* system. Although it is observed that the Vring-HSwheel was rated higher than the V-HSwheel, the results of a GEE model with multinomial distribution suggest that system design was non-significantly related to the colour accuracy rating (Wald $\chi^2(1)$=0.48, p=.487).

The task type had a significant relationship with the colour accuracy rating (Wald $\chi^2(2)$=19.67, $p<.0001$). Queries for T2 task were less frequently assigned a higher score than those for T3 task ($p<.0001$), while queries for T1 task also were less likely to be rated higher than that for T3 task (p=.003). It is estimated that queries for T2 task were more likely to be rated lower than those for T1 task (p=.010), see Figure 7.18.

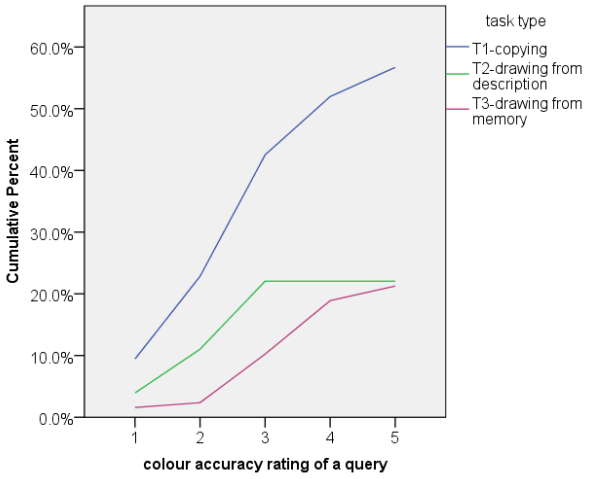


FIGURE 7.18: The observed cumulative colour accuracy ratings of queries drawn via two wheel-based colour pickers.

7.6.2.6 Participants' Usability Ratings

Wilcoxon signed-rank tests show that there were significant differences in the ratings of the relaxation of use, ease of use, and efficiency of colour selection between the two different colour pickers, as demonstrated in Table 7.2. The results indicate that usability rankings on abovementioned features were rated statistically higher for the ***Vring-HSwheel*** System that those for the *V-HSwheel* System. Participants also felt that they could draw **more similar** query to the target painting when using the ***Vring-HSwheel*** System (Median=5) than using the *V-HSWheel* System (Median=3) (Z=-2.59, p=.007).

TABLE 7.2: Results of Wilcoxon tests, Median and IQR of 7-point ratings of usability of *Vring-HSwheel* and *V-HSwheel* colour picker.

Usability features	Significance of Wilcoxon test	Median (IQR)	
		Vring-HSwheel	V-HSwheel
Relaxation	***Z*=-2.56, *p*=.011**	5 (4-6.5)	5 (3-6)
Ease-of-use	***Z*=-2.81, *p*=.005**	5 (4-6)	4 (3-6)
Quick to complete a query	Z=-2.81, p=.099	5 (4-7)	5 (3-6)
Easy to get wanted colour	***Z*=3.00, *p*=.003**	5 (4-6)	4 (3-6)
Quick to pick a colour	Z=-1.63, p=.102	5 (4-6)	5 (3-6)
Enough colour choices	Z=-0.71, p=.479	6 (4-7)	6 (4-7)
Similarity of query	***Z*=-2.59, *p*=.007**	5 (3-5.5)	3 (2-5)

7.6.2.7 Preference and Best-Worst Choices of Colour Picker

A Wilcoxon signed-rank test is conducted to evaluate whether participants' preferences were affected by the shape of the 1D slider. The results reveal that there was a non-significant difference between the preference for colour picker *Vring-HSwheel* and *V(bar)-HSwheel* (Z=-0.03, p=.974), indicating that the shape of 1D slider did not affect participants' preference of wheel-based colour pickers, as shown in Figure 7.19.

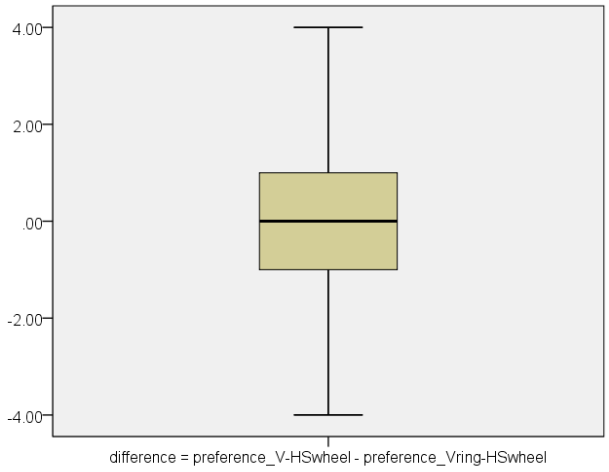


FIGURE 7.19: The symmetry distribution of preference difference between *V-HSwheel* and *Vring-HSwheel*.

A GEE model based on the best-worst choices of wheel-based colour picker suggests that there was not significant difference in participants' best-to-use choices between *V-HSwheel* and *Vring-HSwheel* (Wald $\chi^2(1)$=3.28, p=.070), and the difference in choice of the worst-to-use colour picker was not significant either (Wald $\chi^2(1)$=3.56, p=.059), see Figure 7.20.

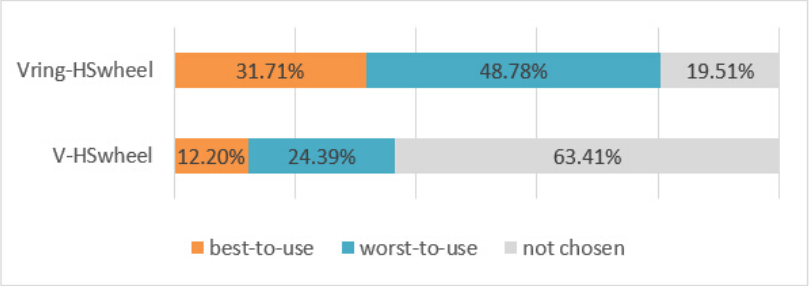


FIGURE 7.20: Proportion of colour picker being selected as the best-to-use and worst-to-use 2D-wheel colour picker.

7.6.3 Shape of 2D input Area - *Square* vs *Wheel*

7.6.3.1 Time of drawing a Query

A GEE model indicates that the colour picker significantly affected the time of drawing a query (Wald $\chi^2(1)=5.15$, $p=.023$), with more time spent on drawing a query using *V-HSsquare* (M=280.47s, SE=25.59) than using ***V-HSwheel*** (M=227.46s, SE=18.50).

There was also a significant effect of task type on query formulation time (Wald $\chi^2(2)=6.26$, $p=.044$), in that participants spent more time drawing a query for T1 task (M=302.70s, SE=13.89) than for T3 task (M=218.86s, SE=34.74) ($p=.020$) and T2 task (M=239.25s, SE=34.52) ($p=.065$), see Figure 7.21. The time difference between T2 and T3 tasks was not significant ($p=.675$).

Drawing skills had a highly significant effect on the total time of query formulation (Wald $\chi^2(3)=17.24$, $p=.001$). Participants with the highest drawing skill (score of 4) spent the least time in drawing a query (M=210.39s, SE=16.71), while participants with moderate drawing skill (drawing score of 3) spent more time (M=282.79s, SE=36.48) than those with drawing scores of 1 and 2, as shown in Figure 7.21.

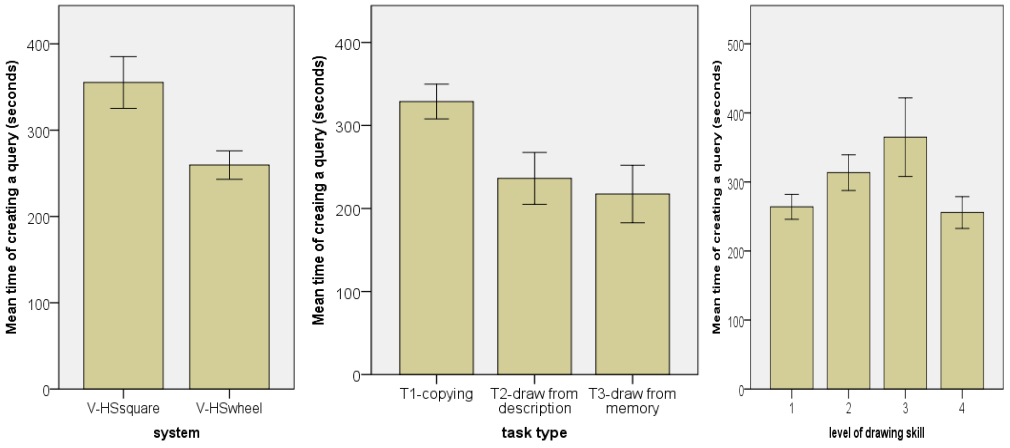


FIGURE 7.21: Main effects of system, task type and drawing skill on the time of drawing a query. Error bars are the standard error of the mean.

7.6.3.2 Number of Colours Depicted in a query

The GEE model analysis reveals that the colour picker design significantly affected the time of drawing a query (Wald $\chi^2(1)$=4.49, p=.034), in that much more colours were depicted in a query when using *V-HSsquare* (M=7.55, SE=0.32) than using *V-HSwheel* (M=7.02, SE=0.30), see Figure 7.22.

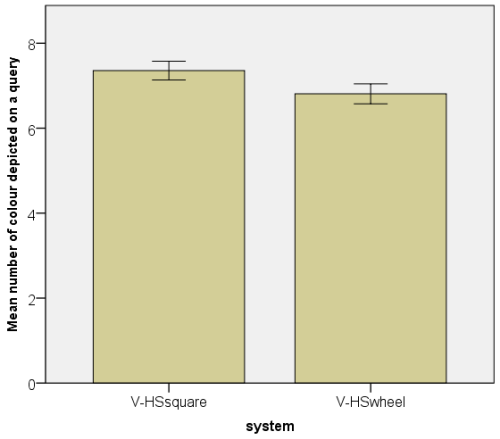


FIGURE 7.22: Main effects of painting type and task type on the number of colours depicted in a query. Error bars are the standard error of the mean.

The task type, painting type and drawing skill did not affect the number of colours drawn in a query.

7.6.3.3 Number of Objects Depicted in a query

A GEE model indicates that there were non-significant effects of system design, painting type, task type, and drawing skill on the number of objects depicted in a query.

7.6.3.4 Search Performance

92 queries are submitted to our server. Overall, 70.65% of queries successfully return the wanted painting in the 51 top results. As shown by Figure 7.23, 72.34% of queries drawn on system ***V-HSwheel*** are successful, compared to 68.89% of queries made on system *V-HSsquare* being Top51-query. All queries for T3 tasks belong to the Top51-query, and more queries produced for T1 tasks (69.01%) are Top51-query than for T2 tasks (54.55%). There are a bit more queries for landscape painting search (72.73%) belonging to Top51-query than for portrait painting (70.37%). Surprisingly, all queries from participants with the highest drawing skill scores are Top51-queries, followed by 75% of queries made by participants with drawing skill scores of 1.

A GEE model reveals that both colour picker (or system) design (Wald $\chi^2(1)=0.37$, $p=.541$) and painting type (Wald $\chi^2(1)=0.17$, $p=.734$) had non-significant effects on the number of Top51 queries.

The GEE results suggest that there was a significant effect of task type on the number of Top51 queries (Wald $\chi^2(2)=251.47$, $p<.0001$). Pairwise comparisons of three task types indicate that more queries produced for T3 tasks were successful compared with those for T1 tasks ($p=.026$) and those for T2 tasks ($p=.006$), regardless of which colour picker was used. The search performance differences between T1 tasks and T2 tasks was non-significant ($p=.763$).

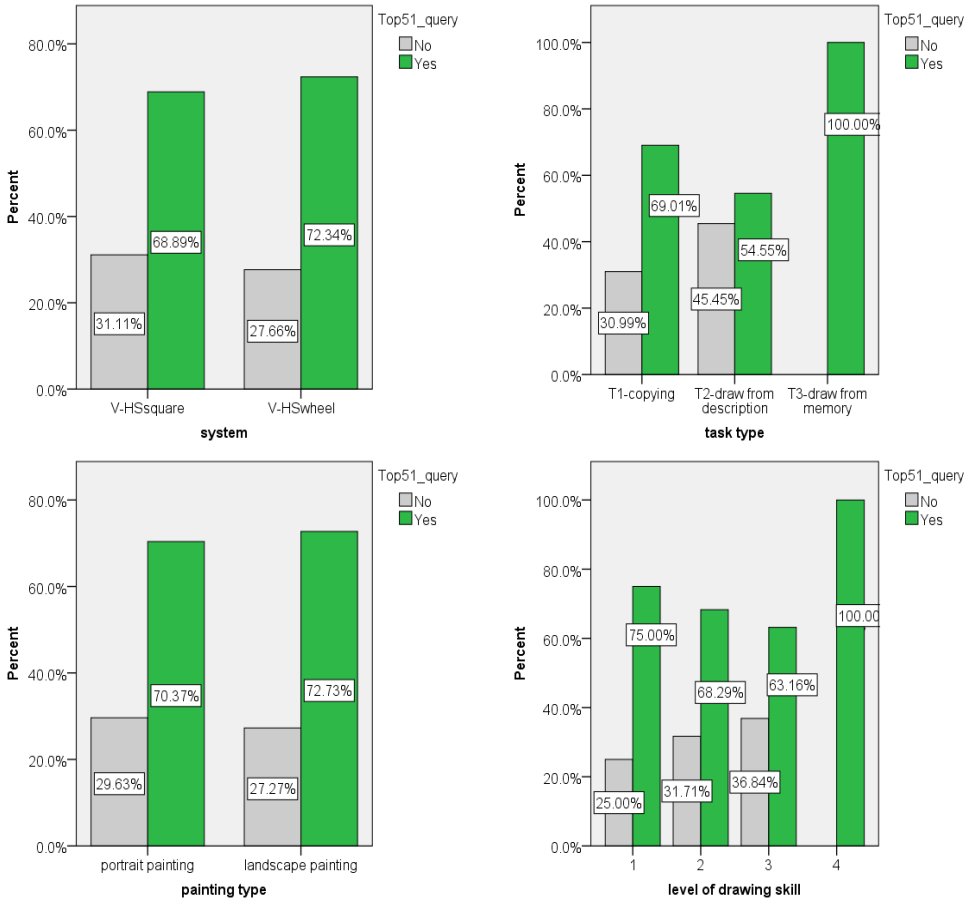


FIGURE 7.23: Percentages of the Top51 query and non-Top51 query in terms of two bar-based colour pickers, task type, painting type, and drawing skill.

7.6.3.5 Human Ratings of Colour Accuracy

Overall, there are 93 queries drawn using system *V-HSsquare* and *V-HSwheel*. The results of a GEE with multinomial distribution model suggest that colour picker design was not related to colour accuracy rating (Wald $\chi^2(1)=0.60$, $p=.438$).

The task type significantly influenced the query's colour accuracy rating (Wald $\chi^2(3)=6.47$, $p=.039$). The parameter estimations indicate that a query for T2 task ($p=.015$) and T1 task ($p=.033$) was less frequently rated higher than a query for **T3** task, and the colour accuracy rating difference between a query for T1 task and a query for T2 task was non-significant ($p=.627$).

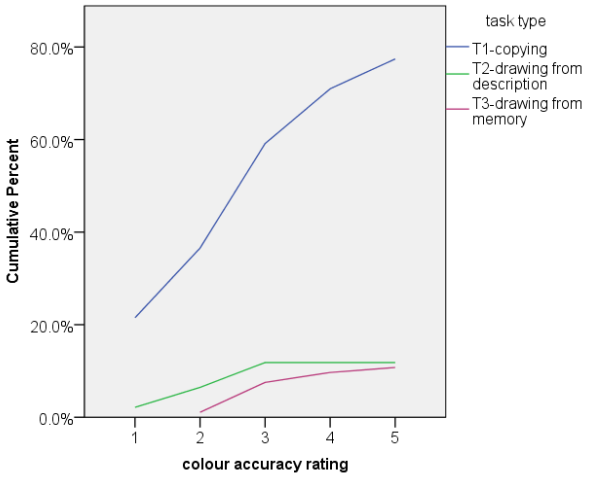


FIGURE 7.24: The observed cumulative colour accuracy ratings of queries made by using *V-HSsquare* and *V-HSwheel*.

7.6.3.6 Participants' Usability Ratings

Wilcoxon signed-rank tests reveal that there were significantly higher ratings on the *relaxation of use* of the ***V-HSwheel*** colour picker (Median=5) than of the *V-HSsquare* colour picker (Median=4) (Z=-2.05, p=.040). Statistically non-significant differences are found in the ratings of other usability features and query similarity between two colour pickers, see Table 7.3.

TABLE 7.3: Results of Wilcoxon tests, Median and IQR of 7-point ratings of usability of *V-HSsquare* and *V-HSwheel* colour picker.

Usability features	Significance of Wilcoxon test	Median (IQR)	
		Vring-HSwheel	V-HSwheel
Relaxation	***Z*=-2.05, *p*=.040**	4 (3-5)	5 (3-6)
Ease-of-use	Z=-0.92, p=.357	4 (3-5)	4 (3-6)
Quick to complete a query	Z=-1.76, p=.078	4 (3-5)	5 (3-6)
Easy to get wanted colour	Z=-1.60, p=.110	4 (3-5)	4 (3-6)
Quick to pick a colour	Z=-1.66, p=.097	4 (2.5-5)	5 (3-6)
Enough colour choices	Z=-1.40, p=.162	5 (4-6)	6 (4-7)
Similarity of query	Z=-0.83, p=.406	4 (2.5-5)	3 (2-5)

7.6.3.7 Preference of Colour Picker

The Wilcoxon signed-rank test is conducted to evaluate whether participants' preferences are affected by colour picker design. The results reveal that a significant difference in preference between *V-HSsquare* colour picker and *V-HSwheel* colour picker (Z=-2.03, p=.001) was found, in that more participants preferred the ***V-HSwheel*** (M=16.86) to the *V-HSsquare* (M=15.7), as shown in Figure 7.25.



FIGURE 7.25: Almost symmetry distribution of preference difference between *V-HSwheel* and *V-HSsquare*.

7.7 Discussion

In Experiment 3, we examine how the three major visual design attributes of a 1D-2D continuous colour picker affect various measures of usability including the colour parameter combination, the visual representation of 1D and 2D input area. This section outlines how each of usability measures is affected by each of the visual design attributes.

7.7.1 Colour Parameter Combination

With the *bar-square* representation of colour picker, three combinations of HSV colour parameters are explored in Experiment 3. The effects of the HSV colour parameter combination are summarised in Table 7.4. The direct effect of colour parameter combination is only found in the number of colours depicted in a query, in that fewer colours were drawn in a query by using the *S-HVsquare* colour picker than using the colour picker with other combinations (H-SV, V-HS).

TABLE 7.4: Summary of the colour parameter combination effects (n.s.: not significant).

Measures	Effects	Significance level
Total time of query formulation	n.s.	p=.534
No. of colours depicted in a query	(H-SV, V-HS) > **S-HV**	***p*=.009**
No. of objects depicted in a query	n.s.	p=.964
Colour accuracy ratings of a query	n.s.	p=.783
Number of Top51 queries	n.s.	p=.319
Preference Choice	n.s.	p=.071
User-rated Best colour picker	**H-SV** > (S-HV, V-HS)	***p*=.019**
User-rated Worst colour picker	n.s.	p=.261
Relaxation	n.s.	p=.359
Ease-of-use	n.s.	p=.746
Easy to get wanted colour	n.s.	p=.229
Quick to pick a wanted colour	n.s.	p=.625
Quick to complete a query	n.s.	p=.865
Enough colour choices	n.s.	p=.086
Similarity of query	n.s.	p=.508

Among all subjective measures from both human evaluators' accuracy ratings and participants' Likert responses on usability, the colour parameter combinations only affect the participant's choices of the best colour picker. It is reported by the participants that the **H-SV** combination was the best colour picker to use among the three colour pickers, but the participant's preference of colour picker did not significantly differ among three combinations.

7.7.2 Shape of 1D Input Area

Two shapes of 1D input area are examined in this experiment: bar and half-ring. Table 7.5 presents the effects of the shape of 1D input area (slider). The shape of slider did not significantly affect the objective measures such as the drawing time, the number of colours and objects drawn in a query, the search performance, neither the subjective outcome measures, including the human evaluator's colour accuracy ratings of a query and the participants' preference and the usability choices.

TABLE 7.5: Summary of effects of the shape of 1D input area (n.s.: not significant).

Measures	Effects	Significance level
Total time of query formulation	n.s.	p=.151
No. of colours depicted in a query	n.s.	p=.534
No. of objects depicted in a query	n.s.	p=.144
Colour accuracy ratings of a query	n.s.	p=.487
Number of Top51 queries	n.s.	p=.692
Preference Choice	n.s.	p=.974
User-rated Best colour picker	n.s.	p=.070
User-rated Worst colour picker	n.s.	p=.059
Relaxation	**Vring-HSwheel** > V-HSwheel	***p*=.011**
Ease-of-use	**Vring-HSwheel** > V-HSwheel	***p*=.005**
Easy to get wanted colour	**Vring-HSwheel** > V-HSwheel	***p*=.003**
Quick to pick a wanted colour	n.s.	p=.102
Quick to complete a query	n.s.	p=.099
Enough colour choices	n.s.	p=.162
Similarity of query	**Vring-HSwheel** > V-HSwheel	***p*=.007**

The only effect of slider shape is found on the participants' subjective usability ratings. In general, the colour picker with ***half-ring*** slider was rated higher than the colour picker with *bar* slider in respect of the relaxation of use, ease of use, and the ease of colour selection. The participants also perceived higher similarity of their drawings when using the ***Vring-HSwheel*** colour picker than using the *V-HSwheel*.

7.7.3 Shape of 2D Input Area

The shape of 2D input area plays an important role during colour selection. Table 7.6 demonstrates the effects of 2D input area on both objective and subjective measures.

TABLE 7.6: Summary of effects of the shape of 2D input area (n.s. =not significant).

Measures	Effects	Significance level
Total time of query formulation	**V-HSsquare** > V-HSwheel	***p*=.023**
No. of colours depicted in a query	**V-HSsquare** > V-HSwheel	***p*=.005**
No. of objects depicted in a query	n.s.	p=.167
Colour accuracy ratings of a query	n.s.	p=.281
Number of Top51 queries	n.s.	p=.541
Preference Choice	**V-HSwheel** > V-HSsquare	***p*=.001**
Relaxation	**V-HSwheel** > V-HSsquare	***p*=.040**
Ease-of-use	n.s.	p=.357
Easy to get wanted colour	n.s.	p=.110
Quick to pick a wanted colour	n.s.	p=.097
Quick to complete a query	n.s.	p=.078
Enough colour choices	n.s.	p=.162
Similarity of query	n.s.	p=.406

The experimental results show that participants spent more time on drawing a query by using the *V-HSsquare* colour picker than using the *V-HSwheel* colour picker, and more colours were drawn in a query when using the *V-HSsquare* colour picker than using the *V-HSwheel*. However, there were non-significant differences in both search performance and query colour accuracy ranking between these two types of colour picker.

Participants reported that the **bar-wheel** colour picker was more preferred to use in comparison with the **bar-square** one, which is in line with the fact that participants perceived the *V-HSwheel* colour picker is more relaxed to use than *V-HSsquare*.

7.8 Summary

This chapter introduces an experiment aimed to explore the relationship between the visual design attributes of 1D-2D continuous colour pickers and the usability of a Query-by-Drawing (QbD) mobile App. The investigated visual design attributes include the combinations of colour parameters, the shape of 1D input area, and the shape of 2D input area.

The experimental results suggest that colour parameter combinations only affect the number of colours depicted in a query and participants' choices of the '*best-to-use*' colour picker. The fewest colours are drawn in a query when using the colour picker with **S-HV** combination, and the *bar-square* colour picker with **H-SV** combination is thought the best colour picker.

The shape of 1D input area only has strong effects on participants ranked usability features, and the colour picker with a *half-ring* slider is perceived better than the colour picker with a bar slider in terms of the relaxation, ease-of-use, and ease-of-colour-selection. Participants also feel that the query drawn by using *half-ring* slider is more accurate than that using *bar* slider.

Consistent effects of the shape of 2D input area are found in both objective measures and subjective measures. It takes more time to complete a query by using the colour picker with *square*-based 2D input area than using the colour picker with *wheel*-based 2D input area. More colours are drawn using the colour square than the colour wheel. Accordingly, the ***wheel*-based** colour picker is perceived more relaxed to use and thus wheel-based colour picker is preferred by participants compared to colour picker with *square*-based input area.

Chapter 8

Experiment 4 - A Longitudinal Memory Drawing Study

8.1 Introduction

There is a quote from Kimon Nicolades [100]: "... *there is no drawing that is not a memory drawing, because no matter how slight the interval is from the time you look at the model until you look at your drawing or painting, you are memorising what you have just seen.*" When users draw a query on a colour-based image search system, it is more likely that people try to remember the colour of an object and then pick the closest colour to compose a query. If the target image is shown, users can draw the query by comparing with the target image; when the target image is previously-viewed but out of sight, users have to draw a query from their memory.

Searching for a previously-seen image is a task that people perform on many occasions. The results of our previous experiments (Ex1 & Ex2) suggest that participants are able to complete all target 30-second present tasks by drawing the spatial composition and the primary colours in the target paintings. However,

it must be noted the task condition under which participants are asked to draw the painting stimulus immediately after 30 second-viewing. In reality, users might be interested to find a painting they have previously viewed for a long while. People's memory might diminish after a long-term delay [3]. Therefore, we design and carry out Experiment 4 to examine whether a user is able to draw the visual information of a previously-viewed painting as a query after three extended retention periods. If the answer is yes, it could be induced that Query-by-Drawing (QbD) mobile applications might be useful for real-world painting search, and accordingly the QbD paradigm would be feasible in the reality. This experiment also reveals what factors and how affect the memory drawing and search performance, e.g., drawing skill and experience with QbD. The drawing pattern and memorisation strategies adopted by participants are also explored. This work has been published in [153].

In this chapter, we first introduce the purposes of Experiment 4 in Section 8.2 and the colour picker interfaces used in this experiment (Section 8.3). The experimental design and procedure are described in Section 8.5). In Section 8.6, we present the statistical analyses of the effects of delayed periods and user characteristics on the task performance and subjective measures (Section 8.6.1). The memorising strategies adopted by participants will be discussed in Section 8.6.2. Participants' drawing patterns will be present in Section 8.6.3 and the general discussions will be given in Section 8.7.

8.2 Aims of Experiment

The primary goal of this experiment is two-fold: to investigate whether people could draw useful queries from memory over several different delay periods, and to have a better understanding of certain attributes affecting the changes of

drawing over time using current colour picker interfaces. This experiment aims to answer the following research question (RQ):

RQ5 : Whether or not people have the ability to draw a query for the painting viewed (15 minutes, 1 week, 1 month) previously? What factors and how affect people's memory drawings of painting over time.

The specific objectives include:

- To explore whether users could memorise and draw the rough memory out as a query input;
- To compare quality of memory drawing and search performance over three time periods;
- To determine whether painting type affects the quality of memory drawing and users' confidence over time;
- To compare whether and how personal characteristics such as gender, drawing skill, and the experience with QbD App affect the memory drawing and users' confidence over time;
- To investigate how users memorise and draw from their memory.

8.3 Colour Picker Interfaces

We conducted the Experiment 4 three months earlier than the Experiment 3, therefore we just choose the colour picker commonly used in current application market rather than the best colour picker tool studied in Experiment 3. The H-SV colour square is used for designing both brush colour picker and background colour picker. Specifically, System B's brush colour picker and System C's

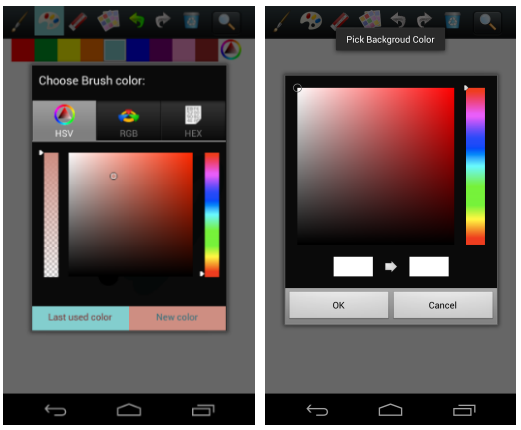


FIGURE 8.1: Screenshots of the brush colour picker (left) and background colour picker (right) of QbD system used for the longitudinal memory drawing study.

background colour picker (please refer to Section 5.3) are used to select the colours for memory drawing, as shown in Figure 8.1.

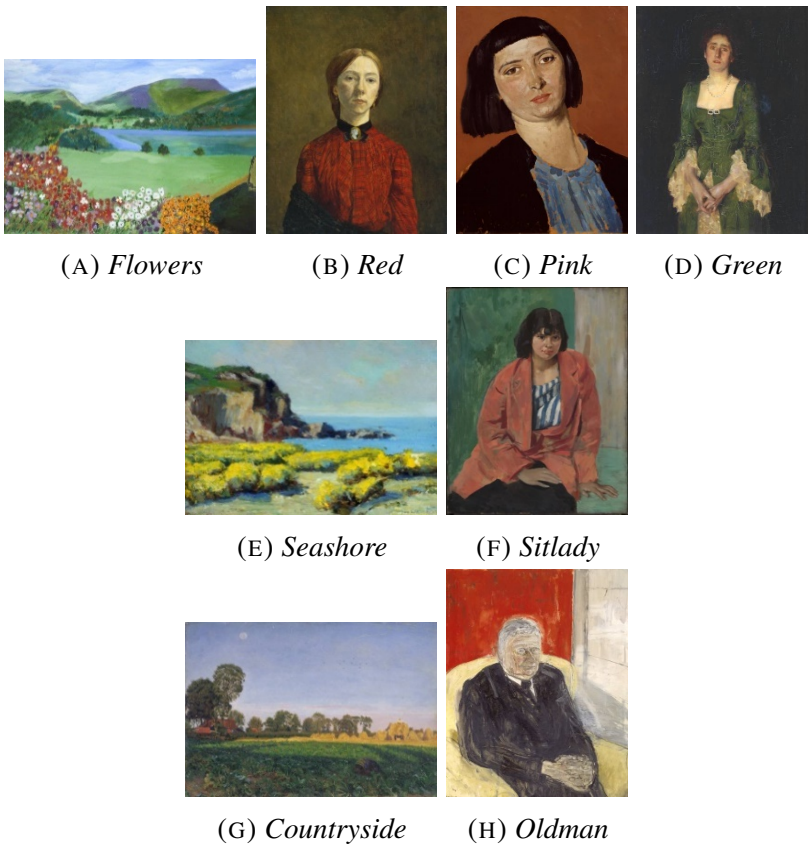
8.4 Database and Result Display

As for Experiment 2, 400 portrait paintings and 400 landscape paintings downloaded from *BBC your paintings* website are used as database images. The 51 top-matched images are shown as results on the users' phone.

8.5 Experimental Design and Procedure

8.5.1 Stimuli

Some works in literature indicate that repeated exposure to stimuli would lead to relatively accurate memory, because there might be a learning effect of the repeated stimuli; while other research demonstrates that multiple exposure to the same stimuli may not necessarily result in enhanced memory [10]. Therefore,



(A) *Flowers* (B) *Red* (C) *Pink* (D) *Green*

(E) *Seashore* (F) *Sitlady*

(G) *Countryside* (H) *Oldman*

FIGURE 8.2: The stimuli with names used in Experiment 4: (A)-(D) are for 15-min delay session; (E)-(F) for 1-week delay; (G)-(H) for 1-month delay).

in our experiment, both singly-shown stimuli and multiply-shown stimuli are used. The stimuli tested at each of three trails include both portrait paintings and landscape paintings with the similar level of complexity, see Figure 8.2.

8.5.2 Participants

17 participants (8 females and 9 males, 7 experienced and 10 novices, age range=22-35) were recruited from the University of Nottingham with an average age of 27.8 years old. Seven experienced participants (2 females, 5 males) were selected from the participants of our previous experiments and are thus familiar with our QbD application. These experienced participants kept their own participant ID used during previous experiment (Participant 4, 5, 8, 9, 15, 28, and

31), while each of the novice participants was assigned a new subject number. The participant ID was recorded along with all drawings and questionnaires.

All participants are right-handed and none self-indicated themselves as drawing-expert or artist. One participant reported having weak sight and was not excluded from the study because participants with colour vision deficits were also our potential users. The examiner obtained the E-mail addresses of participants to remind them of the exact time and place of the second or the third phase of the experiment.

During the participant recruitment, participants were informed that there are three phases of the study lasting for around one month and their participation in each phase is vital to the whole experiment. However, two participants failed to appear for the third phase of 1-month delay due to experiment-unrelated reason.

8.5.3 Procedure

The preparation of the experiment and practice session were the same as for the previous experiments. Following the explanation of the experiment, each participant was given an informed consent form (see Appendix A) to sign. All individual study sessions are one-by-one, with one experimenter and one participant involved. For each query, participants were asked draw as many details as they could. They were told to not worry about the time of drawing each query. After completing each drawing, participants were asked to rate the similarity of their drawings on a 4-point scale, where 1 indicates low similarity. The memorisation strategies adopted by participants were asked in the exit-interview.

During the first study trial (15-minute delay session), participants had 30 seconds to learn each of two paintings shown on the reference phone - the '*Flowers*' and randomised one of the *Red*,*Pink* and *Green*. After that the painting was removed and participants were instructed to finish the pre-questionnaires (see

Appendix A). After 15 minutes of viewing painting, participants were asked to draw their memorised paintings on the phone with screen recorder in a random order. At the end of the 15-minute delayed session, participants were asked to memorise each of the '*Seashore*' and '*Sitlady*' for three minutes, and they were notified that these two paintings would be tested after one week.

After one week, the participants were brought to the study room and were asked to draw and rate their queries for the '*Seashore*' and '*Sitlady*' in random order. We also asked participants to draw the paintings tested during the 15-minute delay trial, which was unexpected by the participants. At the end of the 1-week delay session, the participants were timed to study the '*Countryside*' and '*Oldman*' individually, with each showing for three minutes.

Only 15 participants (7 experienced, 8 novices) returned for the 1-month delay drawing test. Participants were asked to draw and rate their drawings of the '*Countryside*' and '*Oldman*', and afterwards the participants were also instructed to randomly draw two portrait paintings and two landscape paintings tested during the previous two trials.

8.6 Results

The primary goal of this experiment is to characterise the changes of memory drawing over time and to explore the factors that influence these changes. The memorisation process and drawing behaviour are also examined from observations and video-logs.

8.6.1 Statistical Analysis

The outcome measures comprise the time of drawing a query, query completeness (the number of colours and objects depicted in a query), search performance

(the number of Top51 queries), and human ratings of query accuracy. The independent variables include the study trial (coded as 1=15-min delay, 2=1-week delay, 3=1-month delay), painting type (1=portrait painting, 2=landscape painting), the participant's experience with our QbD App (0=novice user, 1=experienced user), gender and drawing skill (self-rated score of 1, 2, 3 out of 5, see Figure 8.3).

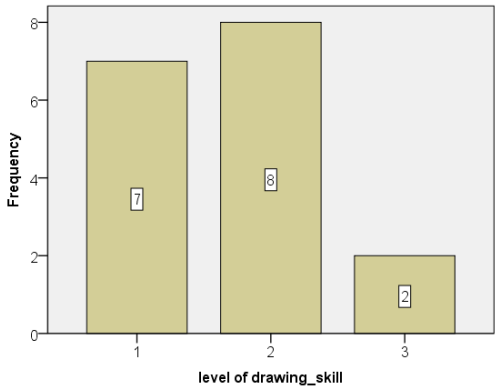


FIGURE 8.3: Distribution of participants' self-rated drawing skill (1=worst, 5=best).

The distributions of the continuous outcome variables are not normal, as revealed by a Shapiro-Wilk test ($p<.01$) and observations of the histogram plots. The outcome measures from the same individuals are collected repeatedly through three time points, therefore these repeated measures from an individual are correlated. Two participants were absent during the 1-month delay session due to study unrelated reasons, thus the response data from three trials is partially incomplete. According to Shek and Ma [117], participant dropout is often encountered in longitudinal studies. The *Generalized Linear Mixed Model* (GLMM) model is more suitable for analysing our longitudinal study with missing data [126], especially in understanding the individual's changes of response from one period to another. The fixed factors of the GLMM model include the study trial, gender, drawing skill, the experience with QbD App, painting type and the interactions among these variables. The random factors contain the stimulus item

and individual participants, because both 8 painting stimuli and 17 participants are sampled from a larger population.

8.6.1.1 Time of drawing a Query

A GLMM model with a gamma distribution suggests that participants spent a significantly different amount of time in drawing a query during three study trials ($F(2,175)=11.07$, $p<.0001$). As shown in Figure 8.4, participants spent much more time making a query during session of 15-min delay (M=339.98s, SE=41.52) than that of 1-week delay (M=269.18s, SE=31.54) ($p=.003$) and that of 1-month delay (M=242.48, SE=28.42) ($p<.001$). The query formulation time difference between last two trials (1-week delay and 1-month delay) was not significant ($p=.084$).

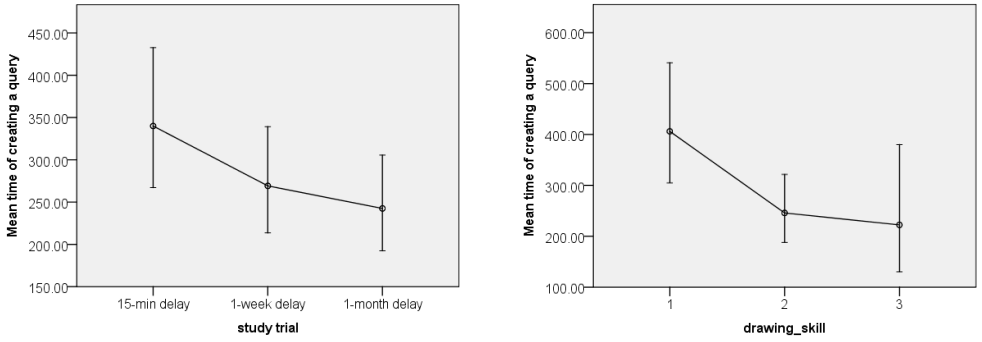


FIGURE 8.4: Main effects of the study trial (left) and drawing skill (right) on the time of drawing a query. Error bars are the 95% confidence interval of the mean.

The results of GLMM also reveal that participant's self-rated drawing skill significantly affected the time of drawing a query ($F(2,175)=3.87$, $p=.023$). Pairwise comparisons indicate that participants with lower drawing skills spent much more time in drawing a query than those with higher drawing skills. Specifically, participants with drawing skill score of 1 (M=406.17s, SE=58.97) spent much more time than those with drawing skill score of 2 (M=245.77s, SE=33.45)

(p=.019) and those with drawing skill score of 3 (M=222.29s, SE=60.43) (p=.031). The query formulation time difference between participants with drawing skill score of 2 and those with drawing skill score of 3 was non-significant (p=.734).

Participant's gender and experience with our QbD App did not affect the time of drawing a query.

8.6.1.2 Number of Colours Depicted in a query

A GLMM model suggests that a significantly different number of colours were drawn in a query between three study trials (F(2,171)=9.35, p<.0001). As the time goes on, participants selected **fewer and fewer colours** to draw a query from memory, see Figure 8.5. The pairwise comparisons reveal that more colours were used for drawing a query during the 15-minute delay trial than the 1-week delay trial (p=.020) and the 1-month delay session (p<.0001), and this difference between the query during the 1-week delay trial and the 1-month delay trial was also significant (p=.007), with fewer colours depicted in a query during the 1-month delay trial than that during the 1-week delay trial.

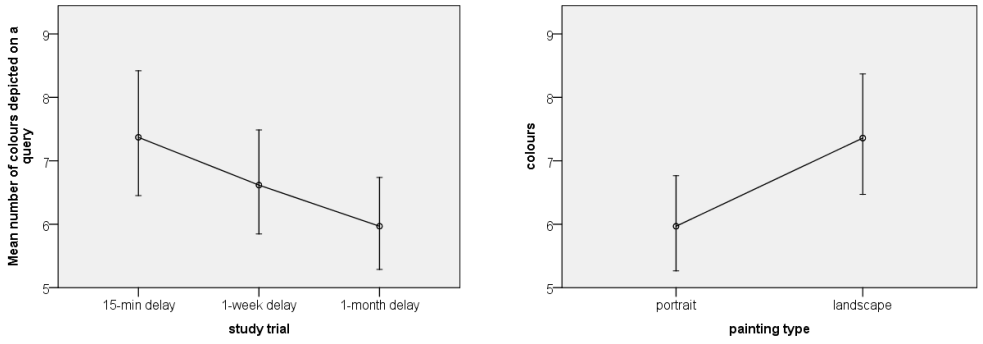


FIGURE 8.5: Main effects of the study trial (left) and painting type (right) on the number of colours depicted in a query. Error bars are the 95% confidence interval of the mean.

The painting type also significantly influenced the number of colours depicted in a query (F(2,171)=14.41, p<.0001). The participants selected much more

colours to draw a landscape painting query (M=7.63, SE=0.39) than a portrait painting query (M=6.16, SE=0.31).

The gender, drawing skill and the experience with using our QbD App did not significantly affect the number of colours used for drawing a query from memory.

8.6.1.3 Number of Objects Depicted in a query

A GLMM model reveals that the number of objects drawn from participants' recollection was decreased as the time passed ($F(2,172)=3.48$, $p=.033$). Pairwise comparisons indicate that participants drew **fewer objects** in a query during **1-month delay** session (M=7.18, SE=0.21) than that at 15-minute delay session (M=7.92, SE=0.30) ($p=.018$) and 1-week delay trial (M=7.60, SE=0.24) ($p=.041$), and while the difference in the number of objects drawn in a query between the first two study trails was not significant ($p=.256$).

It is found that the painting type significantly affected the number of objects ($F(1,172)=9.31$, $p=.003$). **More objects** were depicted in a query for **portrait** painting (M=8.18, SE=0.29) than that for landscape painting (M=6.99, SE=0.27).

The experience with using our QbD App significantly affected the number of objects ($F(1,172)=3.91$, $p=.050$). Pairwise comparison indicates that **novice** participants (M=7.96, SE=0.28) drew **more objects** than those recruited from the previous experiments (M=7.19, SE=0.27). It might because that participants who got familiar with our system would put more focus on the salient (colour) details rather than all objects in a painting.

Other participant characteristics such as gender and drawing skill did not significantly influence the number of objects drawn in a query.

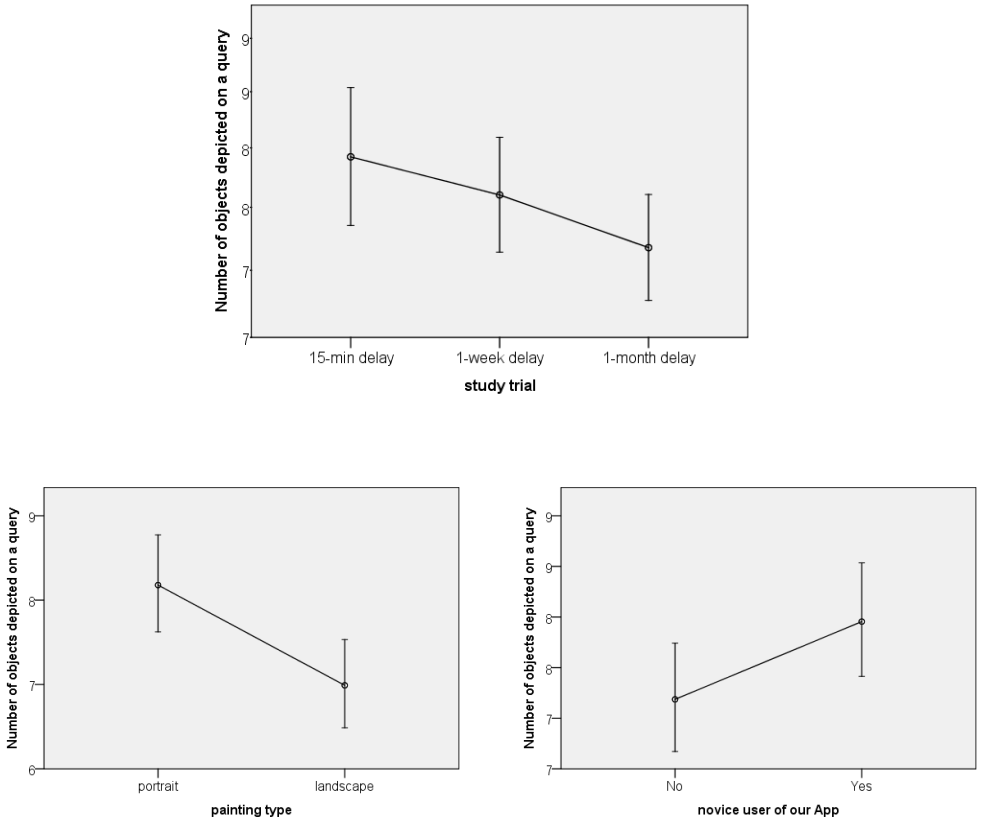


FIGURE 8.6: Main effects of the study trial (top), painting type (bottom-left) and experience with QbD (bottom-right) on the number of objects depicted in a query. Error bars are the 95% confidence intervals of the mean.

8.6.1.4 Search Performance - Top51 Query

The number of Top51 queries is considered as the measure of search performance. Overall, 52.75% of 182 queries from three trials successfully hit the target painting in the returned top 51 images. It is observed that 57.50% of queries made during the trial of 15-minute delay belong to the Top51-query, followed by 55.56% of queries from 1-month delay and 45.90% of 1-week delay queries are the Top51-query, as shown in Figure 8.7. However, the parameter estimates of response variables obtained from a GLMM model reveal that there were non-significant differences in the number of Top51 queries between the three delayed time points ($F(2,177)=0.96$, $p=.384$).

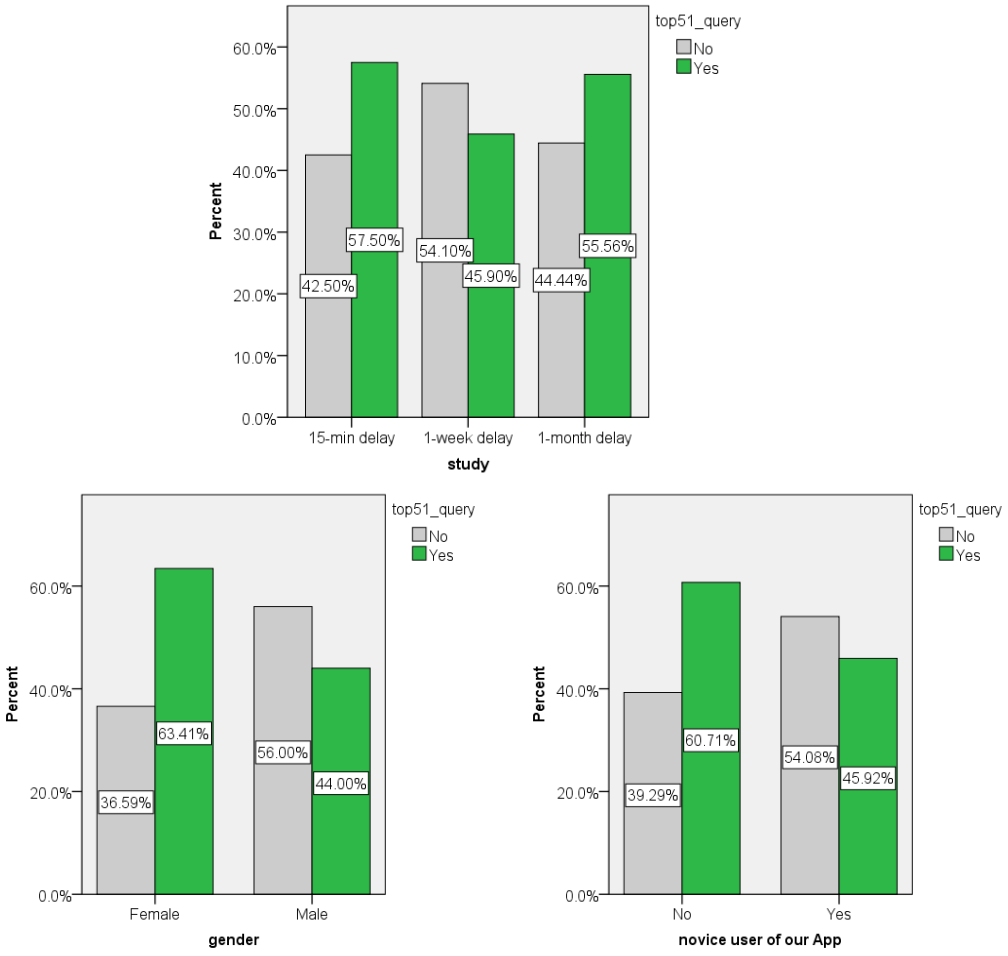


FIGURE 8.7: Proportions of Top51 queries in all queries drawn during three study trials (top), and in terms of gender (bottom-left) and the experience with our QbD App (bottom-right).

We also observe that **females** performed better in drawing more Top51 queries (63.41%) than their male counterparts (44%). The GLMM results validate this observation in that participant's gender significantly affected the number of Top51 queries ($F(1,177)=7.35$, $p=.007$), and the parameter estimations suggest that female participants produced more Top51 queries than males.

The GLMM results also indicate that participant's experience with our QbD App significantly affected the number of Top51 queries ($F(1,177)=6.64$, $p=.011$). The **experienced** participants produced **more Top51 queries** compared to novice users. This is consistent with the descriptive analysis results of the experiment

(see Figure 8.7), in that 60.71% of queries drawn by the experienced participants belong to the Top51 queries, while only 45.92% of queries made by novice participants are Top51 query. The reason might be that participants become more and more familiar with how QbD system works and thus their ability of using our QbD application is improved.

8.6.1.5 Participants' Similarity Ratings

After completing each query, participants were asked to rate on a 4-point Likert scale for the similarity of their drawings comparing to the target painting (1=extremely dissimilar, 4=extremely similar) before submitting the query to the Server. The similarity ranking requires the evaluation of individual drawing against their memory of reference painting based on their own personal standard. Therefore, a high score does not indicate that the drawing is a good one but a high level of confidence of a participant.

Ratings of 175 queries are collected as ordinal data. The overall distribution of the similarity ratings of queries across three trials is shown in Figure 8.8, with more than a half of queries (54.86%) are indicated to be accurate or extremely accurate, and almost half of all queries (46.86%) are rated as accurate (similarity score of 3). The distribution of drawing similarity ratings illustrates that increasingly higher scores of similarity of drawings were rated by participants over time.

This observation is consistent with the results of a GLMM model with multinomial distribution. The GLMM model suggests that the study trial significantly affected the participants' ratings of similarity ($F(2,169)=5.98$, $p=.003$). The parameter estimates indicate that participants were more likely to rate a query from **1-month delay** session as '*extremely similar*' than that query from 15-min delay trial ($p=.002$) and query from 1-week delay trial ($p=.026$). The query from

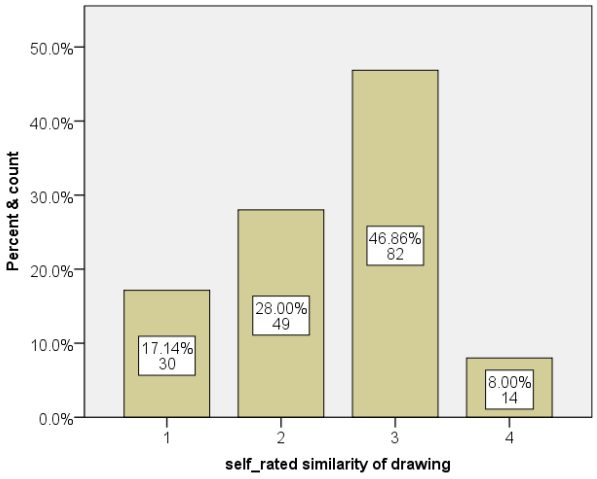


FIGURE 8.8: Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest).

1-week delay study is more likely to be rated as '*extremely similar*' than that from 15-min delay study (p=.201), (see Figure 8.9, left).

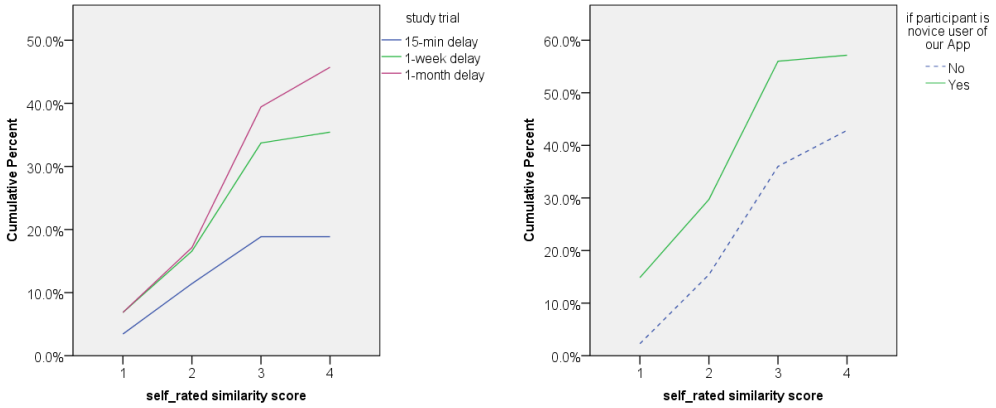


FIGURE 8.9: Plots of cumulative percentage of participants' self-rated similarity ratings regarding study trials (left) and participant's experience with our QbD App (right).

The experience with our QbD App also affected participant's self-rated similarity score (F(1,169)=4.20, p=.042), in that novice participants were more likely to rate a query as '*extremely dissimilar*' than experienced participants. This may signify that participants become more and more confident about both use of our App and similarity ratings of their drawings (see Figure 8.9, right).

Participant's drawing skill and gender, and painting type were not significant predictors, thus being excluded from the GLMM model during model fitting.

8.6.1.6 Evaluators' Accuracy Ratings

182 queries are produced from three study trials over time. The colour accuracy ratings are collected by asking three evaluators to rate on a 5-point scale (1=least accurate, 5=most accurate) about colour accuracy and overall query accuracy of each query.

Colour Accuracy of Query The distribution of colour accuracy ratings is shown in Figure 8.10.

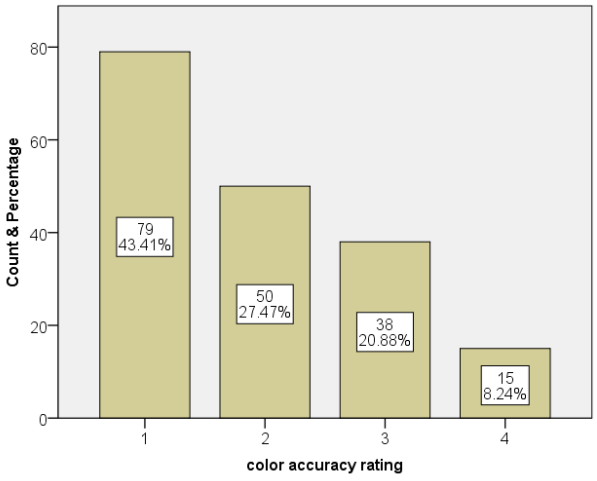


FIGURE 8.10: Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest).

A GLMM model with multinomial distribution reveals that study trial had a highly significant association with colour accuracy ratings ($F(2,173)=5.58$ $p=.004$). As shown in Figure 8.11, the queries from the **15-minute delay** trial were more likely being rated with **higher** colour accuracy scores than those from the 1-month delay session ($p=.001$), while the difference in colour accuracy ranking between queries from the 15-minute and 1-week delay trials was not significant

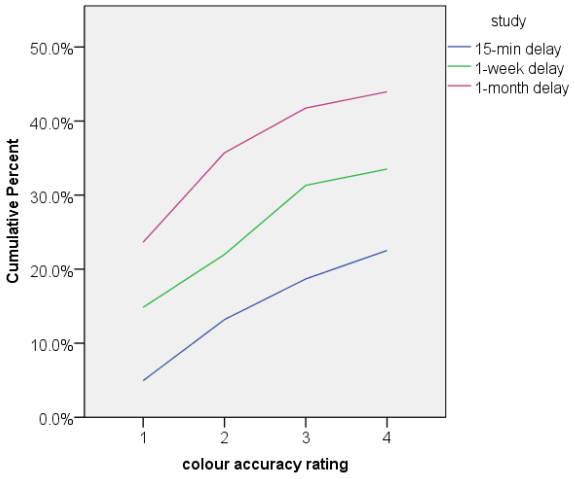


FIGURE 8.11: Plot of cumulative percentage of colour accuracy ratings from three study trials.

(p=.062). Queries from the 1-week delay session were more frequently to be rated higher than those form the 1-month delay study (p=.081). It is indicated that colour accuracy of participant's drawing is decayed as the time passes by.

Other factors such as participant's gender, drawing skill, and experience with our QbD App were not significantly related to human evaluators' colour accuracy ratings.

Accuracy of Overall Query The overall distribution of query accuracy ratings is shown in Figure 8.12. A GLMM model with multinomial distribution was adopted to analyse the query accuracy ratings (5=most accurate). Participant and stimuli are considered as random factors, and participant's novice status, gender, drawing skill and study trial are analysed as fixed effect factors.

The results of GLMM model reveal that study trial was related to query's overall accuracy (F(2,176)=3.33, p=.038). The parameter estimates suggest that queries from the 15-min delay study were more frequently rated with higher accuracy

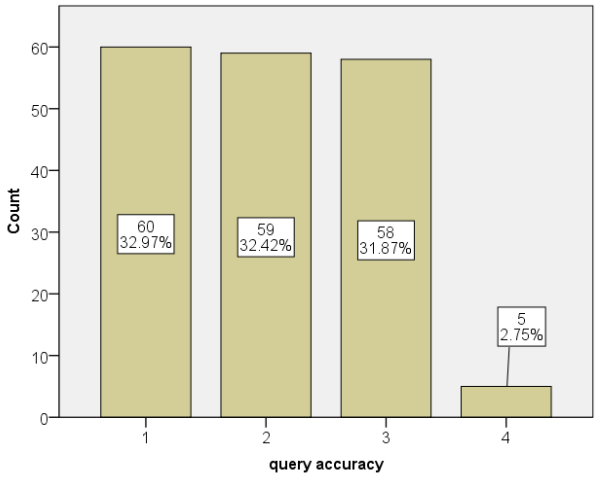


FIGURE 8.12: Distribution of participants' self-rated similarity of their drawings (1=lowest, 4=highest).

scores than those from the 1-month delay study (p=.011), and the accuracy rating of query from 1-week delay trial was likely being rated higher than those from 1-month delay study (p=.216).

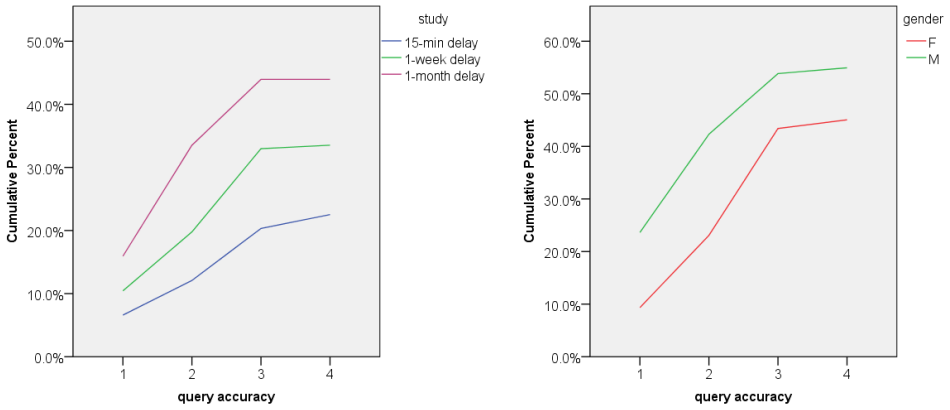


FIGURE 8.13: Plots of the cumulative percentage of participants' self-rated similarity ratings regarding study trials (left) and participant's experience with our QbD App (right).

Both the cumulative percentage plot (see Figure 8.13) and GLMM results indicate that **female** participants' queries were more often assigned higher scores than males' queries ($F(1,175)=3.42$, p=.035).

8.6.2 Memorisation Strategies

Taking together the observations and the screen-recorded drawing activities of participants, it is found that various strategies were adopted by participants to memorise the self-selective details in the target painting during the stimulus studying phase. In line with the literature, the most frequently used strategy is the **gist-based** memory [147]. From participants' post-questionnaires, it is found that the **compositional structure** of a painting is firstly recognised and memorised, followed by the **details** of each sub-part.

The experienced participants realised that making the colours right is more important for better search results. Statistical analyses verify this, in that experienced participants drew fewer objects in a query than novice participants (see Section 8.6.1.3). For example, Participant 4 (expert user, male) reported his memorising strategy: "*I firstly remember the **structure** and then **colours**, taking the* 'Flowers' *as example - there are four layers (in top-to-bottom direction), with top layer is blue sky, then mountain layer, then lake layer, and bottom layer is flowers with red, white, yellow in left-to-right horizontal order.*" The queries of '*Flowers*' from Participant 4 at three delayed time points are shown in Figure 8.14.



FIGURE 8.14: Participant 4's '*Flowers*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial).

Participant 40 (novice user, female) memorised the objects in a painting in a **clockwise composition**. Figure 8.15 shows her drawings from the first two trials. She explained that: "*I first remembered the objects in a clockwise direction*

starting from the bottom flowers, then details such as flowers' colours, green mountain, and blue lake. The primary colour of painting is greenish."

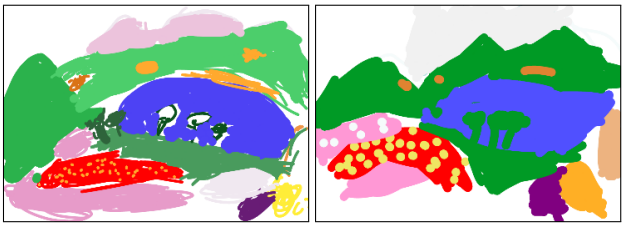


FIGURE 8.15: Participant 40's '*Flowers*' queries during first two study sessions (left-to-right: 15-min delay, and 1-week delay trial).

Some participants simplified the painting memorisation by memorising the **objects with similar colours** together. For example, Participant 8 (experienced, female) and Participant 15 (experienced, male) reported that they memorised the 'sky' together with the 'lake' in the painting '*Flowers*', because both have similar colour. For the painting '*Pink*', Participant 45 (novice, female) memorised one colour for both mouth and background, and one colour for jacket and hair together, see Figure 8.16. But during the 1-month delay session, she confused the jacket colour with the background colour.



FIGURE 8.16: Participant 43's '*Pink*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial).

A **verbal description of the scenes or objects** is adopted by several participants when memorising the target painting. According to [67], language may enhance the symbolic or symbolic-like representation encoding in visual memory. For example, Participant 31 (experienced, female) used her first language to facilitate her memorisation process: "*I remembered two hills and that the middle one*

had purple colours. I told myself (in my native language Germany) the colours are red, white, and orange to remember the sequence of the 'Flowers' *colours. And I also paid attention to the 3 trees.*" Figure 8.17 shows her '*Flowers*' drawings over time.



FIGURE 8.17: Participant 31's '*Flowers*' queries during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial).

Motor encoding (muscle memory) is also used by some participants. For example, Participant 77 (novice, male) traced the painting by moving fingers on the desk or in the air to strengthen his memory.

When the memory of desired painting is blurred and confused, participants probably adopted the inference tactic based on their knowledge and experience. For instance as participant explained "*if there is a land, there might be brown or yellowish.*" More examples will be listed in Section 8.6.3.4.

8.6.3 Drawing Pattern

8.6.3.1 Decayed Memory

The memory decay is represented as the reduced number of objects. For example, as shown in Figure 8.18, fewer objects are depicted on the '*Seashore*' queries after one month delay.

The other indicators of memory decay are drawing wrong details of objects in the query or forgetting all information. Three participants even could not remember the gist-based information in two portrait paintings and one landscape

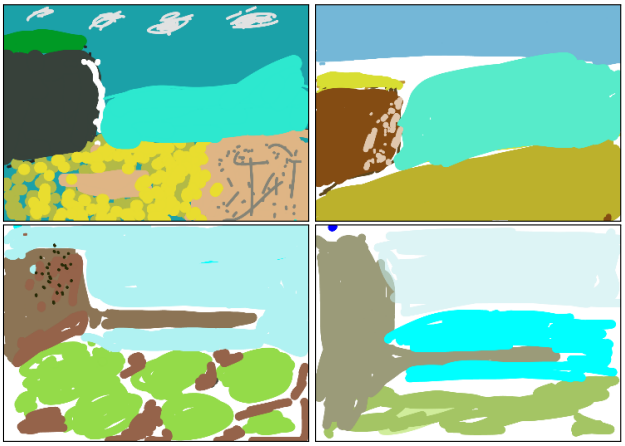


FIGURE 8.18: The '*Seashore*' queries of Participant 28 (experienced, male) (top row) and Participant 81 (novice, female) (bottom row) during 1-week delay (left) and 1-month delay session (right).

painting. Specifically, during the 1-week delay trial, Participant 73 (novice, male) forgot and gave up drawing a query of the '*Flowers*' and Participant 78 (novice, female) could not remember the '*Sitlady*'; Participant 9 (experienced, male) forgot the '*Sitlady*' during the 1-month delay trial. The examiner observed that these participants did not concentrate on the painting during the painting viewing phase.

8.6.3.2 False Memory

It is worthwhile to note that the issue of practice effects in the longitudinal study, even for the wrong information. For example, Participant 81 (novice, female) indicated during the exit-interview that she actually drew from memory of drawing within the previous study trials. As shown in Figure 8.19, the positions of three trees in the queries from the last two sessions are the same but wrong.

Participants' memory of one painting interacts with the memory of other paintings, especially for the landscape painting stimuli. As shown in Figure 8.20, both Participant 75 (novice, female) and Participant 80 (novice, female) drew the '*Seashore*' queries with flowers part of painting '*Flowers*', which indicates



FIGURE 8.19: The '*Flowers*' queries of Participant 81 during three trial sessions (left-to-right: 15-min delay, 1-week delay, and 1-month delay trial).

that participants recollected partial information of one painting as part of another painting.



FIGURE 8.20: The '*Seashore*' queries from Participant 75 during 1-week delay trial (left) and Participant 80 during 1-month delay trial (right).

8.6.3.3 Small Details

The finding of decayed memory is consistent with [10] that people tend to remember gist-based schematic information. They also suggested that details would be forgotten quickly. However, we found that several participants could remember unimportant but attractive details even after the 1-month delay. For example, both Participant 74 (novice, female) and Participant 79 (novice, female) noticed and even drew the 'cottage' and 'pathway' which are very small on the original painting, see Figure 8.21. Six participants (Participant 31, 74, 75, 77, 79, and 81) depicted small trees near the lake through three sessions, an example is shown in Figure 8.21.



FIGURE 8.21: The '*Flowers*' queries of Participant 74 (top row) and Participant 79 (bottom row) during 15-min delay (left), 1-week delay (middle) and 1-month delay (right) session.

8.6.3.4 Experience-based Memory Drawing

The effect of prior experience can be found on the memory drawings. Firstly, remembering inferences is involved in drawing from memory. As shown in Figure 8.22, Participant 76 remembered that the '*Seashore*' depicts a beach scene and he inferred that there must be seagulls, so he drew seagulls on the query during both 1-week delay trial and 1-month delay session, but in fact no any animals are shown in the target painting.

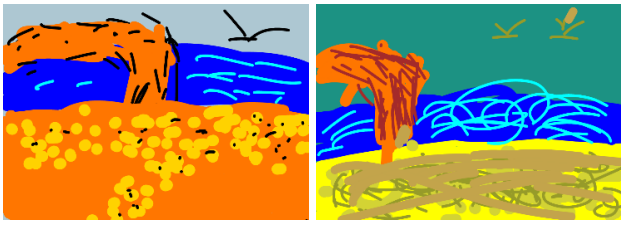


FIGURE 8.22: Queries of the '*Seashore*' from Participant 76 (novice, male) during 1-week delay (left) and 1-month delay (right).

Secondly, a participant's understanding and appreciation of a painting is influenced by his or her previous experience and knowledge. A different understanding of painting would lead to different memory encoding, hence affected

memory drawings. Participants draw quite varied queries for painting '*Countryside*'. For example, Participant 28 (experienced, male) described his drawing as a countryside scene on which a house with white fences are near to the field with haystacks. Participant 76 (novice, male) drew a wheat field view with a human and cottage and he was not sure there was the sun or the moon in the sky. Participant 77 (novice, male) explained his drawing: "... *a pasture with several high trees and haystacks, but I do not know where they were. It might be the sun in the sky.*" See Figure 8.23.

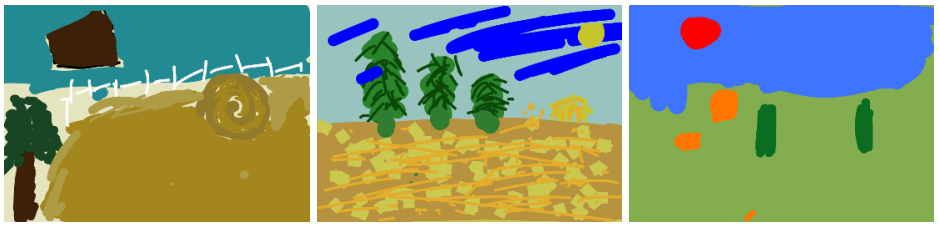


FIGURE 8.23: Queries of the '*Countryside*' from Participant 28 (left), Participant 76 (middle), and Participant 77 (right).

8.7 Discussion

The statistical analyses primarily investigate the effects of delayed periods (study trials) and participant characteristics such as previous experience with the QbD App, gender and drawing skill, on the memory drawing and search performance.

8.7.1 Effects of Delayed Time

The effects of study trials are summarised in Table 8.1. Participants spent more time in drawing a query during the 15-min delay session than during the two later sessions. This is consistent with that more colours were selected during the 15-minute delay trial than the other two trials afterwards, and participants drew the fewest colours and objects during the 1-month delay trial. These findings

are in line with the experimental literature of memory study, in that participants' memory is decayed, e.g., remembering fewer details as the time passing by.

TABLE 8.1: Summary of the delayed trial effects (n.s. = not significant). S1=15min-delay session, S2=1week-delay session, S3=1month-delay session.

Measures	Effects	Significance level
Total time of query formulation	**S1** > (S2, S3)	***p*<.0001**
No. of colours depicted in a query	**S1** > **S2** > **S3**	***p*<.0001**
No. of objects depicted in a query	**S1** > (S2, S3)	***p*=.033**
Number of Top51 queries	n.s.	p=.384
Accuracy ratings of a query	**S1** > (S2, S3)	***p*=.035**
Colour accuracy ratings of a query	**S1** > (S2, S3)	***p*=.004**
Similarity of query	**S3** > (S2, S1)	***p*=.030**

According to accuracy ratings from human evaluators, both colour accuracy and overall accuracy of a query are decreased as time passed, which supports the effect of colour memory degradation proposed by Kok [76]. However, there is a contradiction between participants' self-perceived query similarity and evaluators' judgements on query accuracy and colour accuracy. It is found that participants' similarity ratings increase as the study continued on, which suggests that participants were more and more confident in their memory drawings over time, even though their drawings were actually less and less detailed.

The search results relevant to the number of Top51 queries do not significantly differ over three delayed periods, on average 52.75% of queries drawn from memory successfully found the target painting on first 51 most-matched results. In other words, participants were able to draw a query from their memory and obtained similar search performance after 15-minute, 1-week and 1-month delays.

The statistical analysis results suggest that as the study carried on:

- The 'objective' memory drawing performance, such as the number of Top51 queries, exhibit no obvious decline;

- The 'subjective' memory drawing performance, such as the decreased human evaluators' ratings on colour accuracy and query accuracy, may be impaired.
- The 'subjective' self-perceived similarity score is increased as participants getting more and more familiar with our QbD App.

8.7.2 Effects of Participant Characteristics

The participants in this experiment are divided into groups based on gender, drawing skills and experience with our QbD App.

The effects of gender were mainly on the search performance and human evaluators' query accuracy rating. For both measures, **females** performed better than males, in that females produced more Top51 queries than males, and females' queries were rated more accurate than males' queries by human evaluators. The explanation for these gender differences may be behavioural, because female participants took this experiment more serious than males.

The only direct effect of drawing skills is found on the time spent in drawing a query. Participants with the lowest level of drawing skills spent more time in drawing a query than those with the moderate or higher level of drawing skill.

TABLE 8.2: Summary of significant effects of the experience with QbD App.

Measures	Effects	Significance level
No. of objects depicted in a query	**Novice** > Experienced	***p*=.050**
Number of Top51 queries	**Experienced** > Novice	***p*=.011**
Similarity ratings of query	**Experienced** > Novice	***p*=.042**

There are several effects of participant's experience with our QbD App, as shown in Table 8.2. Experienced participants obtained better search performance than first-time users of our QbD App during all study sessions, and experienced participants were more confident in their own drawings' accuracy than

novices. The results also show that novice participants drew more objects from memory than experienced participants. This might be because some novice users focused on small but 'unimportant' details rather than gist features in a painting to memorise and produce the query, while experienced users paid more attention to the colours of primary blocks in a painting and produced 'better' queries. On the other hand, it can be inferred that the users of our QbD App would get more satisfying search results if they become more familiar with the QbD system.

8.7.3 Effects of Painting Type

The only effect of painting type is found on the completeness of a query. However, the effect of painting type is not consistent between the number of colours and the number of objects depicted in a query. Specifically, participants draw more colours (selecting more colours) for a **landscape** painting query than for portrait painting, while fewer objects are depicted in a landscape painting query than a portrait painting query. The non-significant effect of painting type on the search performance also verifies our selection of stimuli. Several participants reported that portrait paintings are easier to remember than landscape paintings, because landscape paintings are more complicated. However, a larger sample size and more stimuli are needed to validate this. For example, more painting types and more stimuli should be included in each type of painting.

8.8 Summary

This chapter introduces a longitudinal memory-drawing study with real-world painting search tasks (Experiment 4). Three time points after viewing stimuli - 15-minute delay, 1-week delay, and 1-month delay are examined in Experiment

4. The main emphasis is to investigate the participant's ability to recall and to produce a query for painting search by drawing colour sketches from memory over time.

The results show that under the given circumstances, participants depict a query from memory with a decreasing number of colours and objects as time delayed. Although human evaluators' ratings of colour accuracy and query accuracy drop over time, the search performance in terms of the number of Top51 queries does not differ significantly over three trial sessions. Participants became more and more familiar with the system functionality, thus they could complete a query more quickly and also obtain more confidence of their drawings' accuracy. Although novice participants create more objects in a query, experienced participants produce more successful Top51 queries and are more confident than novices.

In general, participants are well able to draw from memory of painting with simple structure (6-7 blobs), even for paintings viewed a month ago. The memorisation strategies and drawing patterns emerging in our experiment verify the gist-based memory in literature. The compositional structure and colour impression of a painting are mainly memorised by participants during the stimuli-studying phase. Memory degradation is found in participants' drawings in terms of decreased and wrong details presented as time passing. The results also provide further evidence that experience and knowledge bias the memory drawing, especially when users forget or their memory blurs.

Chapter 9

Conclusions

A primary goal of this PhD work is to systematically understand and solve the user interface design issues of a Query-by-Drawing (QbD) mobile application. This PhD project contributes to this goal by conducting a series of experiments with three key aspects: 1) a theoretical understanding of how interface elements (e.g., colour picker design and brush type), task type, and user characteristics influence various measures of task performance, usability features and user preference; 2) a basic understanding of how people use and feel about the approach of searching for painting by drawing colour sketches for three types of tasks (from behavioural perspective); and 3) the viability of using the Query-by-Drawing paradigm in real search tasks (e.g., long-term delayed periods).

9.1 Main Contributions

This PhD research has demonstrated the following contributions:

9.1.1 A QbD Prototype as Platform for Future Research

A prototype mobile application of Query-by-Drawing (QbD) painting search is an implementation of ideas and concepts that the thesis author intends to investigate in this PhD project. This QbD application plays an important role as a platform which allows gathering empirical evidence for real user experience. This QbD application adopts the Client-Server architecture, which separates the image search technique from the user interface design, therefore our QbD application could be reusable for testing other CBIR image search algorithms, regardless of which language the QbD algorithm is written by. The user study of examining other interface elements design and psychological experiments are also able to be conducted with our QbD App as a testbed.

9.1.2 Methodology

This thesis presents various novel approaches for conducting a series of drawing-based experiments. Our dedicated methodologies for experimental design, data collection and data analysis are taken from and contribute to the literature.

Firstly, we carried out an online survey to decide the rationales for selecting the stimuli used in our experiments. Secondly, the tasks are designed to simulate realistic problems, which allow participants to draw subjectively based on their own experience as well as objectively tracing the target painting. Thirdly, the methodology used for data collection is novel, such as the screen recorder and the Ziggi USB camera used in our experimental setting (Section 4.4.1). And finally, both descriptive statistics and inferential statistics are adopted for data analysis.

9.1.2.1 Measures

Two primary outcome measures are used in analysing all experimental data: the task performance of using our QbD system (including the time spent in drawing a query, the number of colours and objects in a query, the number of successful queries, and the quality of a query), and participants' subjective attitudes on user interface (including the usability ratings and the preference choices of different interface design).

For the drawing quality, we collect both objective measure - search results, and subjective measure - human evaluators' accuracy ratings. We also define a protocol for assessing drawing accuracy and search result (TopN-query), for example, we asked human evaluators to rate three accuracy ratings for a drawing.

9.1.3 User Interface Design for Colour Picker

The first two experiments lead us to analyse how the user interface design of colour pickers affects the task performance and the usability of a Query-by-Drawing application. Participants generally like the option of having the background colour selection separated from the brush colour selection. Novice users tend to dislike the self-mixed colour picker, while the biggest complaint from expert users is a lack of an advanced drawing brush with effects such as blur, transparency). An eye dropper tool is also desired for some participants.

9.1.3.1 Colour Picker Type - Discrete vs. Continuous

The effects of colour picker type are examined by two groups of colour pickers: one is the brush colour picker group, including two discrete colour pickers and one continuous colour picker; the other is the background colour picker group,

including one discrete colour picker and two continuous colour pickers. It is found that:

- The colour picker type has a non-significant effect on the foreground colour accuracy rating.
- The background colour picked by using colour ring is rated the lowest accuracy score; while for copying tasks, the colour picker type did not affect the background colour accuracy rating.
- Participants perceive that the continuous colour picker provides the most colour choices and the discrete brush colour picker offers the fewest colour choices.
- The continuous colour is chosen by participants as the easiest-to-use colour picker in the brush colour picker group, and the discrete colour picker is chosen as the hardest-to-use background colour picker.
- Most of the participants prefer the continuous colour picker and dislike the discrete colour picker. Participants especially dislike the unordered discrete background colour picker (256 *k-means ordered* database-colours), because it is difficult to select a colour from the disordered representation of colour palette.

9.1.3.2 Combinations of Discrete and Continuous Colour Picker

Each system adopts two colour pickers, one is a foreground colour picker and the other is a background colour picker. Three systems with three '*foreground-background*' colour picker combinations are designed: System A - *discrete-discrete*, System B - *continuous-continuous*, and System C - *discrete-continuous*.

- More objects are depicted in a query when using the *discrete-discrete* System A than other two systems. For copying-tasks, the largest number of colours are used to draw a query on the *discrete-discrete* System A.
- The longest time is taken to produce a query by using the *continuous-continuous* System B, and for copying-tasks, the participants spend the shortest time in drawing a query on the *discrete-continuous* System C.
- System design has non-significant effects on the search performance in terms of the number of queries hitting the target painting within the top 21 most-matched images (refer to the Top21-query) and the number of queries getting the target in the top 51 most-matched images (referred to the Top51-query).
- The *continuous-continuous* System B is chosen by most of 72 participants as the most effort-consuming and the most time-consuming system for all tasks. However, most of 37 participants from Experiment 1 rate System B as the least effort consuming system. This controversial finding is due to the essential biases of the subjective measures [Diener and Suh].

9.1.3.3 Visual Design Attributes of Continuous Colour Picker

Experiment 3 (Chapter 7) examines three key visual design attributes of a continuous colour picker with 1D-2D input areas: 1) the combination of three colour parameters (on HSV colour model); 2) the shape of 1D input area (bar slider vs. half-ring slider); 3) the shape of 2D input area (square vs. wheel).

Colour Parameter Combinations The results suggest that the **H-SV** combination is rated as the best colour picker among three combinations, and participants draw fewer colours in a query when using the colour picker of **S-HV** combination than the other two combinations.

Shape of 1D Input Area The 1D input area is usually represented as a slider. The *bar* slider and the *half-ring* slider are compared in Experiment 3. Only the slider shape only has consistent effects on participants perceived usability rankings. In general, a colour picker with **half-ring slider** is thought better than a colour picker with a bar slider in respect of the *relaxation of use, the ease of use*, and *the efficiency of colour selection*. Participants self-indicated ratings reveal that participants feel more confident when using the **half-ring slider** than using the bar slider.

Shape of 2D Input Area The shape of the 2D input area (*square* vs. *wheel*) plays a more important role in colour selection, as it is highly related to the total time of drawing a query and the number of colours picked to draw a query. The results reveal that the **colour wheel** helps participants to reduce the time of drawing a query compared to the colour square, which is consistent with the finding that participants select more colours via a colour picker with a colour square than with a colour wheel. Qualitative measures indicate that participants prefer the **colour wheel**, and feel **more relaxed** when using the **colour wheel** than using the colour square.

9.1.4 Brush Type Design

It is found that brush type has strong effects on both query formulation and query accuracy. Although more details are drawn in a query using brushes with plain colour, queries drawn using texture brushes are assigned higher score than the queries made using plain colour brush in terms of foreground colour accuracy and overall colour impression. However, the search performance of a query made using plain colour brushes is better than that of a query drawn using texture brushes in terms of the number of Top21 queries. The effect of brush type on the number of Top51 queries is non-significant. Participants spend less time in

drawing a query using texture palettes, due to the limited number of textures. Participants report that texture brushes are easier to draw a query than plain colour brushes, and prefer to use texture brush for landscape painting search.

However, participants usually use texture brushes in the way of first looking through all texture palettes and then selecting one texture that suits their expected colours. The texture self and semantic category are ignored. The preferred feature of texture brushes lies in their realistic look, which is especially superior for a target painting with complex texture.

9.1.5 Task Type and Painting type

Three types of tasks are investigated in this PhD research drawing with the target painting in sight (T1 task), drawing from the description of target painting (T2 task), and drawing the 30-second present target painting (T3 task). Two types of paintings are explored in this PhD project: portrait and landscape paintings.

It is found that participants spend the longest time in completing a query for T1 tasks, because more details are depicted in a query when the target painting is in sight comparing a query for other two types of tasks. Taking all queries from Experiment 1 and Experiment 2, the statistical analyses reveal that the task type has significant effects on both objective and subjective response measures. The task type influences human-rated accuracy of the foreground colour, background colour and colour accuracy of entire query, in that a query for the target-unspecific task (T2 task) is rated more accurate than for the other two types of tasks, and a query for previously-seen painting (T3 task) is assigned the lowest accuracy score of background colour.

Regarding search performance measured by the number of Top21 queries and the number of Top51 queries, the painting type interacts with task type. Queries

for the copying-tasks obtained better search results, with more Top21 queries than queries for 30-second present target paintings (T3 tasks). For all portrait painting searches, participants create more successful queries for target-unspecific tasks (T2 tasks) than for other tasks. For landscape paintings, participants perform better for both copying tasks (T1 tasks) and the tasks of drawing from the description (T2 tasks) than the tasks of drawing from 30-second viewing (T3 tasks). Only for the copying-task and the task of drawing from the description (T2 task), participants perform better for landscape painting search than portrait painting search in terms of the number of Top51 queries, while for the target 30-second present tasks, participants obtain more Top51 queries for portrait painting search than for landscape painting search.

The painting type difference in the number of Top21 queries is non-significant under all types of task conditions. Painting type only influences the number of colours and objects depicted in a query as well as the time of query formulation. A landscape painting query contains more colours and objects than portrait painting query, but somewhat surprisingly, a query for the portrait painting search takes participants more time to complete than landscape painting.

9.1.6 User Characteristics

Participant's gender, ethnic group and drawing skill have also been explored in this PhD project. For drawing-based image search, it is intuitive to wonder whether users' drawing skills affect the task performance. The drawing skills examined in this PhD work is self-indicated ranking (level 1-5) by the participants. It is found that participants' drawing skills only affects the subjective colour accuracy ratings of the query from human evaluators' perspective. In line with Glazek's study [54], queries from participants with higher drawing skills are rated more accurate than those from participants with lower self-rated

drawing skills. However, because our QbD application adopts the colour features, the queries from participants with different levels of drawing skill do not differ significantly in the search performance measures.

Because a western painting database is adopted and the task descriptions are written in English, we also consider participants' ethnic group, including Chinese, European and Others ethnic category. Somewhat surprisingly, the ethnic group does not affect all measures of queries for the copying-tasks and the search performance of queries for all three types of tasks. For the task of drawing a query from the description (T2 tasks), European participants use much more colours to depict a query than Chinese participants; and for the tasks of drawing from short-term memory (T3 tasks), European participants draw more details in a query than people from Other ethnic group (non-European and non-Chinese), therefore queries from European participants are perceived more accurate than those from Others ethnic group.

9.1.7 Drawing and Searching Behaviour

9.1.7.1 Drawing Pattern

The drawing patterns emerge during participants' task performing. Most of the queries are drawn in the same orientation as that of the target painting, i.e., a portrait painting query is made in the portrait screen orientation and a landscape painting query is in landscape screen orientation. Some novice users generate a query by firstly drawing the outlines of objects and then filling colours, same as the way of using pencil-and-paper. After getting familiar with our Query-by-Drawing (QbD) application, participants usually start with producing the background mainly by selecting a background colour or by drawing background objects with brush colours, followed by drawing colour blobs in the canvas. The

default settings of the drawing tools, such as the default size of the 'brush' and the 'eraser', are used more frequently.

With the target painting in sight, participants trace the foreground objects on the reference painting in a systematic sequence (e.g., saliency-based); while for target-absent tasks, right-handed participants usually draw a query in a left-to-right pattern, only one of two left-handed participants draws in a right-to-left fashion. Hence, it is not a good idea to put the tool panel on the left side or the right side of the canvas.

For the target-absent tasks (T2 tasks and T3 tasks), most participants could draw a query from their mental images and they are satisfied with the search results of our QbD application. Given the freedom of using different systems, participants choose a system based on whether it provides the expected colours or textures, and the complexity of selecting wanted colour. In this research, the *discrete-continuous* System C is used more often for drawing a query for the portrait description tasks and the target 30-second present tasks. For the single-system choice, the stimuli do not affect the choice of system. More female participants choose to use the *continuous-continuous* System B than their male counterparts, and more queries for the portrait painting search are drawn on System B than landscape painting search. Participants do not adopt the semantic meaning of the brush textures, but rather mainly consider the colour of each brush texture.

For target-unspecific tasks (T2 tasks), participants usually represent a portrait or scene from the view of the front or side, top. Participants with different background and experience depict different special objects in the queries. And these special objects or colours or the compositional structure depicted in a query or in the participants' head are mainly assessed by participants when judging the relevance of a search result image.

9.1.7.2 Search Strategy

Participants adopt a '*trial and error*' method to search and refine a query and search again. However, this type of search takes longer to find the aimed painting for the users with evolving goals and cautious. Therefore, we propose a three-attempt search condition in our experiment, which may cause another issue that the participants without knowledge of CBIR seem to have no idea what and when is enough for a query to submit for search, because participants realise that they only have three chances available.

During query modification, participants firstly modify the foreground objects, e.g., changing the foreground objects' colour(s), or adding more objects or small details of existing objects. Changing background colour is also firstly concerned by some participants. It seems that most of the participants do not clear which part(s) of a query should be modified for 'better' results.

9.1.8 Viability of Drawing from Memory

Experiment 4 explores memory drawing with three time-delay periods (15-minute delay, 1-week delay, and 1-month delay). The results suggest that the memory indeed decays over time, in that both details of a query and the time spent in drawing a query are decreased gradually over time, but participants become more and more confident on the accuracy of their memory drawings. The tested time points do not affect the search performance (the number of Top51 queries). On average, more than half of the queries successfully return the target painting within the top 51 most-matched search results. Both evaluator-rated colour accuracy and overall accuracy of a query decrease as the time passes, which is in line with the effect of colour memory degradation proposed by Kok [76].

There are several effects of user characteristics. Female participants produce more successful Top51 queries than males, which is consistent with the evaluators' query accuracy ratings. Participants with the lowest drawing skills spend more time drawing a query than those with higher drawing skill. The experience with our QbD mobile App also has strong effects on the number of objects drawn in a query, the number of Top51 queries and self-rated query similarity. Novice participants depict more objects in a query, however, experienced participants draw more Top51 queries and are more confident of the similarity of their drawings than novices. It is worthwhile to note that some novice participants even remember and draw the small but 'unimportant' objects in the target painting, while experienced participants mainly focus on the colours of primary blocks and create a rough query.

In general, participants are able to draw from the memory of a painting with simple compositional structure (with 6-7 colour blobs), even for paintings viewed a month ago. Participants firstly memorise the gist information in a painting, e.g., the primary colour impression and the compositional structure, etc. Several participants adopt a verbal description (in their mother tongue) or the motion memory to facilitate their memorisations. Some participants indicate that landscape paintings are easier to remember than portrait paintings. However, a larger number of samples and stimuli are needed to validate these. Our experiment provides further evidence that experience and knowledge bias participants' memory drawings, especially when participants forget or their memory is faded and uncertain. Participants' drawings from the previous study session might also affect their drawings during the later session, because the same mistakes are found in individual participant's queries for a stimulus painting through three trials.

9.1.9 Design Best Practice

There are a few guidelines and heuristics intended to aid a designer to design a drawing application in the literature, especially for choosing colour pickers with thousands of visual design. Some general guidelines are derived from our observations and participants' feedbacks on effective colour picker design:

- Brushes with five sizes and erasers with five sizes are sufficient to use, and the size of the eraser should be larger than that of the brush.
- Users may have difficulty in drawing some small objects, suggesting that the canvas should have the ability of zooming-in and zooming-out.
- More feedback mechanisms should be implemented, such as showing users the view of the selected colour with actual use, the feedback from matching results.
- It is helpful to include an eyedropper and the pre-defined geometric shapes.
- It is better not to put the drawing tool panel on the left or the right side of the screen of mobile devices, because of the commonly-used '*left-to-right*' drawing sequence.
- The texture brush is helpful for the user to draw a query roughly.

9.2 General Limitations and Future Work

9.2.1 Extensions of Prototype Application

9.2.1.1 Image Matching Techniques

At present, our QbD App only focuses on the matching with colour features by using *Fast Multi-resolution Image Querying* (FMIQ) algorithm [69]. It somehow needs to consider the spatial relationship between colour blobs, e.g., as in *Angular Radial Partitioning* (ARP) [20]. Other algorithms that enable the real-time response and the higher precision and recall of image search techniques could be explored on our QbD App in the future. The image matching with the black-and-white line drawings could be another potential direction for extended research.

9.2.1.2 User Interface

There are a number of features that could be included in the user interfaces of drawing tools, but have been excluded to make our experiments reliable, such as maintaining colour strokes over multiple layers, providing special brush type, e.g., marker and ink brushes, and pre-processing users' drawings before submitting to search (e.g., the old-fashion reproduction). Several topics related to the interface design have been summarised during the data analysis of experiments, but not explored in this PhD project.

Like many colour selection tools, the colour pickers in this PhD research appear as dialogue boxes covering the whole screen and obstructing the view of the very objects to which colour will be applied, and participants can only guess what the colour will look like in its actual usage. In future, we will explore colour

selection tools that allow users to view colours in the context where colours will be used.

Because novice participants tend to draw in the way of using pencil-on-the-paper, colour-filling tools would be helpful. We actually have already explored the *flood-fill*[1] method for automatically filling colours. However, due to the limited time of this PhD project, the effective real-time colour-filling or texture-filling algorithm have not been implemented by the thesis author yet, leading to future further research. Pre-defined geometric shapes and the multi-layer functionality are also considered to be added into the next version of a QbD application.

9.2.2 Support for Database Images

The current database used in experiments separates the portrait paintings from the landscape paintings, which works well for our prototype implementation. Obviously, the real world painting database consists of a large number of different types of paintings, as well as the paintings with various issues, such as the blurredly scanned copy of the painting, line sketches, and overlapped paintings, etc. Therefore, in a real use environment, more images and painting types should be included and tested as the database.

9.2.3 Further Experiments

Although our experiments are conducted in a meeting room or office, which is not a serious experimental environment. Participants are asked to trace or memorise the painting stimuli presented on one Android Phone with the same setting as the other phone on which participants used to draw. This means that both the

[1] https://en.wikipedia.org/wiki/Flood_fill

illumination and the small size of phone screen may influence the participants' appreciation of the art pieces. It would be interesting to see the task performance when allowing users to directly produce the query for the painting hung on the wall in front of the users, e.g., in museums or art galleries. The comparison of using small-screen phones and tablets would be another potential direction. The other types of paintings and a larger size of the samples and stimuli will be included in future experiments, especially for the longitudinal memory drawing study.

9.3 Discussion

The primary objective of this PhD work is to generate an empirical foundation for designing a drawing-based system of search painting on mobile devices. Four key topics are investigated in this thesis: 1) the user interface design of colour picker and brush type; 2) the effects of task type and user characteristics on the usability and task performance of a Query-by-Drawing system; 3) the viability of drawing from memory after several long-term delays; 4) the users' drawing behaviour and attitudes on using Query-by-Drawing application for real-world tasks.

Most participants think that the conception of search painting by drawing a visual query is intuitive and amazing, which is similar as querying music by rhythm [22]. Participants generally prefer the option of separating the background colour selection from the brush colour selection. The self-mixed continuous colour picker is the most preferred colour picker, while the biggest complaint from the expert users is a lack of more advanced drawing brush (such as blur, transparency). An eye dropper is also desired. We cannot hope to match the performance of every user with the current version of the QbD system. However, we believe that our initial experiences with the users and our preliminary

efforts in understanding the design factors have provided the foundation for creating the user-friendly QbD system.

Overall, this research has achieved the objectives of this PhD project that exploring the factors affecting Query-by-Drawing mobile application based on current image retrieval technique, specifically the user interface design aspects such as colour picker interfaces and brush type. The overall success of this work is the empirical exploration into the factors affecting a Query-by-Drawing application, which could be useful in generating the guidelines for mobile Query-by-Drawing application design.

Appendix A

Experiment Forms

A.1 Consent Form

Informed Consent form

Name of Study: To explore the factors affecting visual query formulation of 'A is for Art' Mobile App
Name of Researcher: Min Zhang
Name of Supervisor: Dr. Sarah Atkinson, Dr. Natasha Alechina, Dr. Guoping Qiu

This study is to explore if colour and visual memory affect user's drawing a query on a painting-related Mobile App 'A is for Art'. You will be asked to perform a set of tasks individually on an Android-Mobile phone with the phone screen-recorded and video-camera recording, which will take no longer than 1 hour. The tasks involve drawing several objects on the phone and completing questionnaires. A short interview (discussion) will be taken immediately after performing the tasks to get your experience of the tasks.

	Please tick to agree
I have been invited to participate in the above study, and I confirm that I have read and fully understood the information sheet.	☐
I have had opportunity to ask any questions about it.	☐
I consent to take part in the above study and agree to be video recorded during the study.	☐
I understand that I am free to withdraw at any time without giving any reason.	☐

This is to confirm that I have agreed to take part in this research study on the

date: ……

Name of participant (Please print): ……

Signature of participant: ……

In addition to the data analysis, I give permission for data that could identify me (e.g. photos, video, audio) to be published ☐

Signed: ……..

A.2 T2 Task Descriptions

T2 Tasks: Help Tom to find the painting as he described below:

Tom:
*"Can anyone help me? I'm looking for a **half-length portrait** painting I saw a year ago at an art museum. My memory is blurry…*

a. It was a portrait of a blond lady in yellow coat shown in bluish-green background.

b. Find a portrait of person with brown hair and pink clothes on yellowish-green background.

c. Find a girl with purple hat and purple coat, blond hair, and white collar. She holds an unfolded blue book in her hands on bluish-grey background.

d. Find a man in a light blue suit and a dark blue tie. He is in front of greenish curtain.

Find a countryside scene with sky, mountain, river, tree, and land in:

- Spring,
- Summer,
- Autumn,
- Winter

A.3 Pre-Questionnaire

Pre-task Questionnaire

Age: _____ **Colour-blind** ☐No ☐Yes (please specify)_____

Gender: ☐Male ☐Female **First language:** _____

Your current Occupation (in which area): _____

Please specify major of your current study_____

1. Do you have previous experience of free-hand drawing, e.g., , etc.)?

☐pen-and-paper drawing

☐finger drawing on the mobile phone – DrawSomething

☐mouse/digital pen drawing on the Computer

☐Others e.g. _____

☐Not at all

2. How would you rate your own drawing skills?

Very Bad ☐☐☐☐☐ Excellent

3. How much experience do you have with the following activities, please tick:

How much experience do you have with:	**1 –** Not at all Experienced	**2 –** Slightly Experienced	**3 –** Somewhat Experienced	**4 –** Moderately Experienced	**5 –** Extremely Experienced
a. **- search engines, e.g. Google**					
b. **- image search engines, e.g. Google Images.**					
c. **- drawing your queries on PC or devices to find an object**					
d. **- free-hand drawing**					
e. **- drawing on a computer using mouse**					
f. **- drawing on touch-screen tablet**					
g. **- drawing on touch-screen mobile phone**					
h. **- drawing on touch-screen mobile devices using fingers**					
i. **- drawing on touch-screen mobile devices using pen/stylus**					
j. **- browsing artworks on mobile devices**					
k. **- image management system, e.g. Flickr**					

4. How often do you do the following activities, please tick

How often do you do the following:	Daily	Weekly	Monthly	More seldom	Never
- draw on the paper/canvas					
- draw on computer using mouse					
- draw on mobile device using fingers					
- draw on mobile device using pen/stylus					

5. Do you have any prior experience with visual query systems, e.g.

☐Google Goggle,

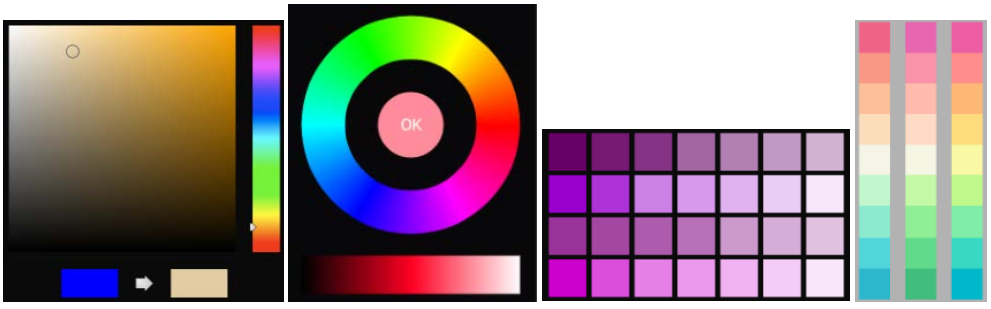
☐Taking a photo and search on the system,

☐Any content-based image retrieval system, e.g. search by colour

☐Other image search engine, e.g. _____

6. If you have any education in visual community, fine arts or similar studies beyond school, please describe below (College, degree, post-graduate degree)

7. Which kinds of colour palette have you used before? Tick them all:



☐ ☐ ☐ ☐

A.4 Post-Questionnaire Q3

1) Compare system ABC and choose the one that:

	System A	System B	System C	
you spent **least effort** on drawing the query (✔)				Spent **Most** effort on drawing the query (✘)
you spent **less time** to get the wanted result (✔)				you spent **most time** to get the wanted result (✘)
Easiest to pick wanted brush colour (✔)				**Most difficult** to pick wanted brush colour (✘)
Easiest to pick wanted background colour(✔)				**Most difficult** to pick wanted background colour(✘)
Has **enough** choice of brush colours (✔)				Has **least** choice of brush colours (✘)
Has **enough** choice of background colours (✔)				Has **least** choice of background colours (✘)
you **like most** to pick brush colour (✔)				you **dislike most** to pick brush colour (✘)
you **like most** to pick background colour (✔)				you **dislike most** to pick background colour (✘)

- ❖ Pick brush colour:



System A



System B



System C

- ❖ pick background colour:



System A



System B



System C

2) Which interface do you prefer to user?

- For Portrait search: ☐Pure colour brush ☐ Pattern brush
- For Landscape search: ☐Pure colour brush ☐ Pattern brush

3) Which brush is easier to use?

- For Portrait search: ☐Pure colour brush ☐ Pattern brush
- For Landscape search: ☐Pure colour brush ☐ Pattern brush

A.5 Post-Questionnaire Q5

2) Was the Mobile App ___ ?

	Extremely	Quite	Slightly	Neither	Slightly	Quite	Extremely	
Satisfying								Frustrating
Reliable								Unreliable
Simple to use								Complex to use
Novel								Standard
Stimulating								Dull
Fast								Slow
Useful								Useless

3) Which features of the system(s) do you prefer to use? And why?

4) Which system(s) do you dislike and what functions would you expect to see?

Appendix B

Colour Values and Order of K-means Colour Picker

```
{ "611f1c", "ad831d", "72653c", "594d41", "d08430", "687d8d", "2a4060" },
{ "c82247", "6f6a66", "435467", "a36b49", "6a5b2d", "159993", "6b541c" },
{ "9d574f", "d0c2a4", "836d3c", "616c79", "492d27", "b7a273", "252e2f" },
{ "53554d", "252113", "b9ac47", "98854d", "dfc690", "a2745f", "a05e37" },
{ "afb8a6", "c9a569", "24231d", "e3c8a5", "232124", "cbcdd7", "cbb892" },
{ "382d1d", "f0a45a", "b38c6b", "532a21", "444440", "632136", "6f200b" },
{ "90897a", "302717", "98442a", "b1a994", "94885d", "7e7360", "4d4d36" },
{ "a0876f", "716e57", "332422", "782321", "b4803e", "d6caae", "5e6c63" },
{ "dfdfdb", "393030", "c28e47", "b9805b", "da8d10", "f4e6c7", "ddd3b9" },
{ "634325", "a6a69f", "3e2b26", "453119", "595779", "c27d0d", "8d6c50" },
{ "6e5557", "c5e1ee", "e9eaec", "d3ad8a", "d08868", "3d2610", "7b0319" },
{ "8f7746", "857d54", "9bafc1", "798b97", "3a3826", "806231", "274c80" },
{ "efc06a", "c0932c", "677b7c", "964741", "6e91b6", "7c6f4a", "c8b8c4" },
{ "afbdc7", "f1d3b6", "96a0a1", "933808", "594c19", "837f66", "fdec90" },
{ "858175", "2b7871", "8496a5", "946f28", "e7b67f", "2b282b", "dfac1c" },
{ "d09e57", "1d1b15", "6e7967", "141513", "873130", "5f594e", "613819" },
{ "3a3934", "553a23", "aa8e4d", "9b9491", "f0c391", "030407", "989c4e" },
{ "90211a", "723522", "52332f", "232828", "1d2324", "fdfdfc", "d69c3f" },
{ "927c6a", "2f2e30", "c09769", "9fabaf", "bdae97", "181a1a", "8ba4b8" },
{ "1c1d2a", "585c39", "181f1f", "4b0e11", "dbb999", "1e1e1d", "68a7d6" },
{ "c0c5c2", "342d2a", "f0d9de", "ceb1a2", "111c27", "513711", "7e6554" },
{ "797673", "ae9679", "2c3a43", "20121d", "f7e8af", "7b8682", "3b454d" },
{ "89550c", "44221f", "8a5e44", "deaf55", "21303a", "723f36", "ac8481" },
{ "a3956c", "4c4b46", "92765a", "2a1c1b", "dabfbe", "ea4c3b", "0e161a" },
{ "476453", "1d130d", "4f4030", "b6b7b7", "936a3b", "e1d9c8", "b3a286" },
{ "565b5d", "8b9a92", "a8c3b7", "303536", "af712c", "2d69ac", "2c190f" },
{ "19282d", "4c4522", "292621", "7c6320", "484e55", "b3aca8", "7b5d44" },
{ "a0793a", "5e3e33", "521d10", "0e584f", "a87b4e", "71430c", "2e2d25" },
{ "c39b7c", "487285", "ad9a5d", "5393c9", "51656d", "bf4c2c", "393d3e" },
{ "6a4f47", "6a4a30", "bd9458", "c1ba9e", "a59b84", "f3d5a0", "745438" },
{ "ac311c", "bd684f", "6f604c", "d0cfc3", "82401f", "f7d37f", "0b1012" },
{ "c1ab89", "424231", "f6f0dc", "b79896", "c8ad7b", "395646", "665941" },
{ "ef7c4e", "9b927e", "aabdda", "320e07", "c6c5b5", "52423c", "a5865f" },
{ "634e39", "d7bc81", "666359", "8b8c8a", "5a4c2e", "704924", "ddaf6b" },
{ "a16b0d", "61684c", "b75d06", "8c5733", "918b6d", "5c7da8", "7d4f2a" },
{ "622f26", "391c1a", "443430", "d6a17a", "c1b8ad", "a5a091", "120c08" },
{ "935c1f", "473a39", "473925", "83463e" }*/ };
```

FIGURE B.1: Colour Value for k-means Background Colour Palette used in Experiment 1

```
"6e604e", "e39d67", "7d6435", "241b0f", "281e19", "b59970", "a78b67",
"9da6a0", "2b2c33", "472306", "a5291f", "2f5d6b", "e5b335", "c2baa8",
"898c87", "9c9684", "4a2628", "9c6445", "4e4143", "495367", "3c424a",
"ab9f80", "5897ac", "47432a", "2c1f23", "57646d", "7f4b09", "468384",
"382b17", "682011", "6f2a20", "86711c", "b8b19a", "48352a", "304d2b",
"9a6122", "1d1e1d", "2c3952", "ce887c", "110704", "a09673", "997c66",
"cec8bb", "5c1d23", "403a23", "dabf88", "5d5a4d", "775147", "170f05",
"1d130d", "c96644", "e61e10", "ae6336", "68978a", "636559", "ed544e",
"acc3d1", "ba3d36", "05090d", "108bba", "1a1f0c", "444c52", "e4d5c0",
"807560", "935c09", "e1cbaf", "200904", "a9a093", "77521d", "07212c",
"916934", "110f13", "9b3b32", "2d554d", "4f371e", "565539", "481104",
"b9bdb9", "211717", "694a24", "243a72", "14150e", "330b04", "d5b56a",
"bb802f", "0a2049", "b25255", "b1985a", "82200f", "2b2311", "3e1a11",
"938a77", "e3e2d7", "9c5034", "656c6a", "384215", "809ea5", "515053",
"1c2226", "8f6c71", "971608", "ae700c", "080904", "b0e5f3", "f3daae",
"63777c", "a28555", "c8a29a", "353120", "cfdfe8", "6c4538", "191a14",
"427462", "ac4019", "894423", "743d28", "40423c", "381905", "99afb7",
"ad7647", "242520", "eeb685", "5a2407", "342218", "785c65", "25272b",
"bb8866", "c89b67", "343435", "ae867c", "876b44", "5d7194", "41231d",
"626242", "302e2a", "072e65", "3f3109", "095764", "511c11", "645341",
"15171a", "c8b285", "8d7f57", "2c281b", "7a976d", "1e2537", "65515c",
"8ba785", "705b3c", "d2953c", "adb0ab", "bb864e", "726c45", "948c66",
"0a7684", "502918", "7d6651", "d4d5cc", "875829", "8a9fca", "70c3d8",
"201605", "b5a98b", "32382b", "ce3917", "343c3f", "625032", "3b312b",
"3a455f", "071418", "86816a", "623b20", "3d3c31", "453216", "674512",
"592e25", "4a3e36", "5e8766", "a16f58", "474c43", "02592e", "d2825a",
"e3482b", "844038", "495d4b", "2b2b23", "533d2b", "969795", "0d0f09",
"758597", "5e3016", "36240c", "dfba9e", "202215", "6b5b2b", "60422d",
"ceb697", "cd8d12", "564f45", "020202", "957946", "52360a", "f1f0ee",
"b17565", "15424c", "2a3141", "412f21", "775030", "d1c3a8", "862f28",
"1b3d1a", "5e0f04", "4d4a35", "073539", "b00608", "d8a582", "ebcd97",
"524219", "7e7f7d", "b42608", "6e3d4a", "073811", "41343a", "761105",
"2b1d07", "8e6b56", "653332", "86583d", "ebc76f", "2f1811", "c8a751",
"161528", "bf4f29", "263110", "584638", "362823", "933412", "f6e6cb",
"442611", "a1923f", "1f3736", "485f7f", "291307", "c1aa75", "6e3409",
"76756e", "3c652f", "c09b83", "7c3519", "9c7d2c", "574c28", "5a7648",
"7b7a4f", "90524e", "706d5d", "545e5c" }
```

FIGURE B.2: Colour Value for k-means Background Colour Palette for portrait painting search in Experiment 2

```
"8f9059", "d8decc", "808378", "2b764d", "4d423a", "b1c3ce", "586a2b",
  "6a7033", "f3c22b", "7ab6d8", "a9aeaa", "57210c", "6b5c25", "80452a",
  "7591ab", "d19d13", "2a4568", "4d868f", "5f5a35", "72a3c7", "ddc9b4",
  "37362e", "d4e7df", "496757", "b06e2c", "202525", "b3c6bd", "cad5de",
  "d6dbdb", "d1c9bb", "7c2f12", "985b38", "a9aa82", "64829e", "acbbb5",
  "c4b8a9", "c98652", "c3ceb9", "be936e", "b6b6aa", "4a462a", "4d6477",
  "3e4237", "e8c79b", "3b5743", "d6d3d0", "b8c5ae", "241c08", "dbd1c3",
  "a4b7c4", "6e734c", "a08f7f", "648e90", "61bcec", "825a2d", "877465",
  "f3daa8", "5d797b", "72883c", "c6d67a", "908e83", "dba13e", "9dafa4",
  "f3f6f2", "0c4f44", "a2a853", "72695f", "5e5b16", "bad1c9", "573c2c",
  "959293", "7eb5c0", "b6aca0", "495236", "85788a", "938473", "2b4b36",
  "bb9f82", "087ca5", "816f50", "d1d6c3", "40291c", "c9cec5", "555042",
  "3a5379", "889971", "302e0e", "d66929", "866e33", "4a391c", "a7a29c",
  "cad7d0", "9a7f62", "e6d0ab", "93b3d7", "634439", "9fb7b3", "191918",
  "bed3ee", "796c24", "6f977f", "436c8c", "d8d6b5", "8a8425", "a09d8e",
  "1b3111", "d0b79d", "44494d", "a18033", "b6a392", "b6cadf", "dbe9f1",
  "59471d", "e9d586", "328ba4", "32751f", "716139", "73604d", "a7c5ee",
  "d7c9a5", "c7c39e", "c9c7af", "b1965e", "587b47", "a7d2da", "767b5e",
  "616543", "cfc7cc", "c5a057", "826342", "a98a12", "acae97", "90adaf",
  "38391e", "9da8b7", "292218", "7b8898", "b9ad40", "543b0a", "093332",
  "bd872e", "865315", "b1a8b8", "795340", "6a9658", "c7c1ba", "066765",
  "a4bed8", "b48f44", "2b6c7e", "caaa8c", "745f0d", "9ba178", "a46511",
  "8fabc5", "d8c691", "858666", "956f41", "6e856e", "23451c", "ad7046",
  "47816e", "381c09", "2e5756", "6f7d86", "5a6066", "bbebf2", "f2e6c7",
  "ebd6ba", "b9b489", "947753", "828250", "95c2ca", "6c4722", "515725",
  "c1ccd4", "68a7af", "bbdedf", "5b6e8f", "fbf447", "656975", "7a733e",
  "e3b785", "ae7e5a", "ae9e72", "899e8d", "ab4322", "0da3c6", "95b988",
  "28332f", "a29053", "989fa8", "e0e2e1", "8b9943", "bdbcb4", "4e6140",
  "739c99", "adbda1", "5397a5", "6a7c1a", "8a730e", "d0a36e", "7f8c88",
  "c1b299", "bfab78", "ad9785", "617562", "956a26", "e9cf5c", "6e5033",
  "078c83", "d1ba78", "555751", "9e8348", "ebe6d8", "494b12", "97a599",
  "052219", "4e72a8", "41a3ce", "f3f4dd", "747671", "d2bfac", "3c4925",
  "8fd3e9", "070a07", "5b4c2f", "e4b068", "46372b", "b6b9c2", "c9bc8b",
  "675546", "171008", "618dbc", "2f2d22", "bfc5c6", "cfddeb", "8b803e",
  "e4dcc5", "88959f", "67480c", "9c9467", "a98a6e", "60311e", "616555",
  "cbbf57", "2c3947", "b6b266", "e3ded3", "455464", "8e6351", "869fb5",
  "40350a", "a49834", "eaece9", "84a7a2" } };
```

FIGURE B.3: Colour Value for k-means Background Colour Palette for landscape painting search in Experiment 2

Bibliography

[1] Albers, J. (2006). *Interaction of Color*. Yale University Press.

[2] Albert, W. and Tullis, T. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Newnes.

[3] Arnheim, R. (2004). *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press.

[4] Atwood, M. E. and Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, 8(2):191–216.

[5] Bailey, P., Manktelow, K., and Olomolaiye, P. (2003). Examination of the color selection process within digital design for the built environment. *Proceedings of Theory and Practice of Computer Graphics, 2003.*

[6] Banfi, F. (2000). *Content-Based Image Retrieval Using Hand-Drawn Sketches and Local Features: a Study on Visual Dissimilarity*. PhD thesis, Institut d'Informatique Universite de Fribourg.

[7] Bartleson, C. J. (1960). Memory Colors of Familiar Objects. *Journal of the Optical Society of America*, 50(1):73–77.

[8] Bauersfeld, P. F., Slater, J. L., Group, A. H. I., Apple, and Letraset (1991). User-oriented color interface design: direct manipulation of color in context. *SIGCHI Conference on Human Factors in Computing Systems*, pages 417–418.

[9] Berlin, B. and Kay, P. (1969). *Basic Color Terms: their Universality and Evolution*. University of California Press, Berkeley and Los Angeles.

[10] Blake, A. B., Nazarian, M., and Castel, A. D. (2015). The Apple of the mind's eye: Everyday attention, metamemory, and reconstructive memory for the Apple logo. *The Quarterly Journal of Experimental Psychology*, (April):1–8.

[11] Blaser, A. (2000). *Sketching spatial queries*. PhD thesis, University of Maine.

[12] Bogles, J. (1989). *Drawing with and without models : An examination of drawing behavior in children and adults*. PhD thesis, M.A thesis, Wilfrid Laurier University.

[13] Brewer, C. A. (1999). Color use guidelines for data representation. In *Proceedings of the Section on Statistical Graphics, American Statistical Association,*, pages 55–60.

[14] Brewer, C. A. (2005). *Designing Better Maps: A Guide for GIS Users*. ESRI Press.

[15] Broek, E. L. V. D. (2005). *Human-Centered Content-Based Image Retrieval*. PhD thesis, Radboud University Nijmegen.

[16] Budd, C. W. (2007). Content Based Image Retrieval Using Sketched Input. Master's thesis, University of Bath.

[17] Cao, Y., Wang, H., Wang, C., Li, Z., Zhang, L., and Zhang, L. (2010). MindFinder: interactive sketch-based image search on millions of images. In *Proceedings of the international conference on Multimedia - MM '10*, pages 1605–1608, New York, New York, USA. ACM Press.

[18] Carson, L. C. (2012). *Drawing Accuracy, Quality and Expertise*. Phd thesis, University of Waterloo.

[19] Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., Stephenson, K., Siegler, R., and Keating, D. (1996). The role of central conceptual structures in the development of children's scientific and mathematical thought. *Monographs of the Society for Research in Child Development*, 61(1/2):1–295.

[20] Chalechale, A., Mertins, A., and Naghdy, G. (2004). Edge image description using angular radial partitioning. *IEE Proceedings - Vision, Image, and Signal Processing*, 151(2):93–101.

[21] Chan, H. C. and Wang, Y. (2004). Human factors in color-based image retrieval: an empirical study on size estimate accuracies. *Journal of Visual Communication and Image Representation*, 15(2):113–131.

[22] Chen, J. and Chen, A. (1998). Query by rhythm: an approach for song retrieval in music databases. In *Proceedings Eighth International Workshop on Research Issues in Data Engineering. Continuous-Media Databases and Applications*, pages 139–146. IEEE Comput. Soc.

[23] Cohen, D. J. (2005). Look little, look often: the influence of gaze frequency on drawing accuracy. *Perception & psychophysics*, 67(6):997–1009.

[24] Cohen, D. J. . and Earls, H. (2008). How shape constancy relates to drawing accuracy. *Psychology of Aesthetics, Creativity, and the Art*, 2(1):8–19.

[25] Cohen, D. J. and Bennett, S. (1997). Why can't most people draw what they see? *Journal of experimental psychology. Human perception and performance*, 23(3):609–21.

[26] Cohen, D. J. and Earls, H. (2010). Inverting an image does not improve drawing accuracy. *Psychology of Aesthetics, Creativity, and the Arts*, 4(3):168–172.

[27] Corridoni, J., Bimbo, a. D., and Pala, P. (1998). Retrieval of paintings using effects induced by color features. In *Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Database*, pages 2–11.

[28] Cunningham, S., Bainbridge, D., and Masoodian, M. (2004). How people describe their image information needs: a grounded theory analysis of visual arts queries. *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004.*, pages 47–48.

[29] D'Ath, P. J., David Thomson, W., and Wilkins, A. J. (2007). Memory for the color of non-monochromatic lights. *Color Research & Application*, 32(1):11–15.

[30] Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval:Ideas, Influences, and Trends of the New Age.

[31] de Graft Acquah, H. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Development and Agricultural Economics*, 2(1):1–6.

[32] de Sciascio, E. and Guaragnella, C. (2000). Color fragmentation-weighted histogram for sketch based image queries. *Proc. EUSIPCO Signal*.

[33] De Silva, G. C., Yamasaki, T., and Aizawa, K. (2009). Sketch-based spatial queries for the retrieval of human locomotion patterns in smart environments. *Advances in Multimedia*, 2009:1–13.

[34] Del Bimbo, A. (1999). *Visual information retrieval*. Morgan Kaufmann Publishers Inc.

[35] Del Bimbo, A. and Pala, P. (1997). Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):121–132.

[36] Derefeldt, G., Swartling, T., Berggrund, U., and Bodrogi, P. (2004). Cognitive color. *Color Research & Application*, 29(1):7–19.

[Diener and Suh] Diener, E. and Suh, E. Measuring Quality of Life: Economic, Social, and Subjective Indicators. *Social Indicators Research*, 40(1-2):189–216.

[38] Douglas, S. and Kirkpatrick, T. (1996). Do color models really make a difference? pages 399–405.

[39] Douglas, S. A. and Kirkpatrick, A. E. (1999). Model and representation: the effect of visual feedback on human performance in a color picker interface. *ACM Transactions on Graphics*, 18(2):96–127.

[40] Dunlop, M. (2000). Reflections on Mira: interactive evaluation in information retrieval. *American Society for Information Science*, 51:1269–1274.

[41] Dupplaw, D., Lewis, P., and Dobie, M. (1999). Spatial Colour Matching for Content Based Retrieval and Navigation.

[42] Durrett, H. J. and Stimmel, D. T. (1987). Color and the instructional use of the computer. In Durrett, H. J., editor, *Color and the Computer*, pages 241–253. Academic Press Professional, Inc.

[43] Ellis, G. and Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. In *BEyond time and errors novel evaluation methods for information visualization - BELIV '06*, pages 1–7, New York, New York, USA. ACM Press.

[44] Epps, H. H. and Kaya, N. (2004). Color matching from memory. In *AIC 2004 Color and Paints, Interim Meeting of the International Color Association*, pages 18–21.

[45] Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., and Equitz, W. (1994). Efficient and effective Querying by Image Content. *Journal of Intelligent Information Systems*, 3(3-4):231–262.

[46] Foley, J. D., van Dam, A., Feiner, S. K., and Hughes, J. F. (1990). Computer graphics: principles and practice (2nd ed.).

[47] Fonseca, M. J. (2004). *Sketch-Based Retrieval in Large Sets of Drawings*. Phd thesis, UNIVERSIDADE T ECNICA DE LISBOA INSTITUTO SUPERIOR TECNICO.

[48] Ford, A. and Roberts, A. (1998). Colour Space Conversions. Technical report.

[49] Francis, M. A. and Irwin, R. J. (1998). Stability of memory for colour in context. *Memory (Hove, England)*, 6(6):609–21.

[50] Gainotti, G., Silveri, M. C., Villa, G., and Caltagirone, C. (1983). Drawing objects from memory in aphasia. *Brain : a journal of neurology*, 106 (Pt 3):613–622.

[51] George Ebbinason, A. and Rajesh Kanna, B. (2014). ColorFingers : Improved Multi-Touch Color Picker. In *SIGGRAPH Asia 2014 Technical Briefs*, pages 1–4. ACM.

[52] Giangreco, I. (2010). Content-based Image Retrieval using User-drawn Color Sketches.

[53] Giangreco, I. (2013). *ADAM A Database and Information Retrieval System for Storing , Organizing*. PhD thesis, University of Basel.

[54] Glazek, K. (2012). Visual and motor processing in visual artists: Implications for cognitive and neural mechanisms. *Psychology of Aesthetics, Creativity, and the Arts*, 6(2):155–167.

[55] Goodnow, J. (1977). *Children's Drawing*. Billings & Sons Ltd.

[56] Goodrum, A. a., Goodrum, A. a., Bejune, M. M., Bejune, M. M., Siochi, A. C., and Siochi, A. C. (2003). A State Transition Analysis of Image Search Patterns on the Web. *Image and Video Retrieval*, pages 193–197.

[57] Granier, X. and Heidrich, W. (2003). A simple layered RGB BRDF model. *Graphical Models*, 65(4):171–184.

[58] Granzier, J. J. M., Brenner, E., and Smeets, J. B. J. (2009). Do people match surface reflectance fundamentally differently than they match emitted light? *Vision research*, 49(7):702–7.

[59] Hamwi, V. and Landis, C. (1955). Memory for Color. *The Journal of Psychology*, 39(1):183–194.

[60] Hanley, J. A., Negassa, A., deB Edwardes, M. D., and Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology*, 157(4):364–75.

[61] Hayashi, T., Ishikawa, A., and Onai, R. (2007). Landscape Image Retrieval with Query by Sketch and Icon. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 11(1):61–70.

[62] Heimonen, T. (2012). *Design and Evaluation of User Interfaces for Mobile Web Search*. PhD thesis, University of Tampere.

[63] Henry, P., Cheung, V., and Westland, S. (2008). Accuracy of Cross-Media Colour Memory. *International Colour Association AIC*, pages 1–4.

[64] Henry, P., Westland, S., and Cheung, T. (2006). An intuitive color-selection tool. In *Color Imaging Conference*, pages 144–147. Society for Imaging Science and Technology.

[65] Hove, L. (2007). Evaluating use of interfaces for visual query specification. *Proceedings of NOBOKIT*.

[66] Hove, L.-j. (2010). *Drawing Visual Query Images-use, users and usability of query by drawing interfaces for content based image retrieval system*. Phd thesis, University of Bergen.

[67] Huette, S., Kello, C. T., Rhodes, T., and Spivey, M. J. (2013). Drawing from memory: hand-eye coordination at multiple scales. *PloS one*, 8(3):e58464.

[68] Hurlbert, A. C. and Ling, Y. (2006). Contextual effects of familiar object colours on colour perception. In *Perception ECVP abstract*, volume 35. Pion Ltd.

[69] Jacobs, C. E., Finkelstein, A., and Salesin, D. H. (1995). Fast multiresolution image querying. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*, pages 277–286, New York, New York, USA. ACM Press.

[70] Järvelin, K. (2009). Explaining user performance in information retrieval: Challenges to IR evaluation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5766 LNCS(1):289–296.

[71] Johnson, P. (2003). *Art: A New History*. Weidenfeld & Nicolson.

[72] Kaiser, P. K. and Boynton, R. M. (1996). *Human color vision*. Optical Society of America.

[73] Kaufmann, R. and O'Neill, M. C. (1993). Colour names and focal colours on electronic displays. *Ergonomics*.

[74] Kherfi, M. L., Ziou, D., and Bernardi, A. (2004). Image Retrieval from the World Wide Web. *ACM Computing Surveys*, 36(1):35–67.

[75] Kim, S., Tak, Y., Nam, Y., and Hwang, E. (2005). mCLOVER: mobile content-based leaf image retrieval system. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 215–216.

[76] Kok, T. (2008). SUIT- Scientic user interface testing. Master's thesis, Radboud Universiteit Nijmegen.

[77] Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. Grune & Stratton.

[78] Korfhage, R. R. (1997). *Information Storage and Retrieval*. Wiley.

[79] Kozbelt, A. and Serafin, J. (2009). Dynamic Evaluation of High- and Low-Creativity Drawings by Artist and Nonartist Raters. *Creativity Research Journal*, 21(4):349–360.

[80] Lai, T.-s. (2000). *CHROMA : A Photographic Image Retrieval System*. PhD thesis, University of Sunderland.

[81] Lai, T.-S. and Tait, J. (1999). CHROMA (demonstration abstract). In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99*, number February, page 324.

[82] Laws, E. (1999). *An Investigation of Color Memory as a Function of Hue, Saturation, Lightness and Observer Imagery Vividness for Blue, Green and Orange Test Hues.* PhD thesis, Virginia Polytechnic Institute and State University.

[83] Levine, D. S. and Perlovsky, L. I. (2008). Simplying Heuristics versus Careful Thinking: Scientific Analysis of Millenniall Spiritual issues. *Zygon*, 43(4):797–821.

[84] Lew, M. S. (2000). Next-generation Web searches for visual content. *Computer*, 33(11):46–53.

[85] Lezak, M. D. (2004). *Neuropsychological Assessment*. Oxford University Press.

[86] Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

[87] Ling, Y. and Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, 25(6):1215–1226.

[88] Ly, M. H. and Hardeberg, J. Y. (2013). Introducing Modern Memory Color. In *AIC2013*, number 1.

[89] Lyons, P., Moretti, G., and Paul Lyons, G. M. (2004). Nine Tools for Generating Harmonious Colour Schemes. *Proc 6th Asia Pacific Conference on Computer Human Interaction (APCHI 2004)*, 2004:1–10.

[90] Lyons, P., Moretti, G., and Wilson, M. (2000). Colour group selection for computer interfaces Incorporated. *Proc Conf Human Vision and Electronic Imaging (SPIE2000)*, 3959:302–313.

[91] McCamy, C. S. (1998). On the number of discernible colors. *Color Research & Application*, 23(5):337–337.

[92] McDonald, S., Lai, T.-S., and Tait, J. (2001). Evaluating a content based image retrieval system. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '01*, pages 232–240.

[93] Mcdonald, S. and Tait, J. (2003). Search Strategies in Content-Based Image Retrieval. In *ACM SIGIR '03*, pages 80–87.

[94] McManus, I., Chamberlain, R., Loo, P., Riley, H., Rankin, Q., and Brunswick, N. (2010). Art Students Who Cannot Draw: Exploring the Relations Between Drawing Ability, Visual Memory, Accuracy of Copying, and Dyslexia.

[95] Mohammad, Y. F. O. and Nishida, T. (2007). NaturalDraw: Interacive Perception Based Drawing for Everyone. In *Proceedings of the 12th international conference on Intelligent user interfaces - IUI '07*, pages 28–31, New York, New York, USA. ACM Press.

[96] Moretti, G. S. (2010). *A calculation of colours : towards the automatic creation of graphical user interface colour schemes*. PhD thesis, Massey University.

[97] Munsell, A. H. (1905). *A Color Notation A measured color system, based on the three qualities Hue, Value and Chroma [Kindle Edition]*. George H. Ellis.

[98] Murch, G. M. (1984). *The Effective Use of Color: Cognitive Principles*. TEKniques.

[99] Newhall, S. M., Burnham, R. W., and Clark, J. R. (1957). Comparison of successive with simultaneous color matching: Errata.

[100] Nicolaïdes, K. (1990). *The Natural Way to Draw: A Working Plan for Art Study*. Houghton Mifflin Harcourt.

[101] Nielsen, J. (1994). Heuristic evaluation. In *Usability inspection methods*, pages 25–62. John Wiley & Sons, Inc.

[102] Nilsson, T. H. and Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of experimental psychology. Human perception and performance*, 7(1):141–150.

[103] Ostrofsky, J., Kozbelt, A., and Seidel, A. (2012). Perceptual constancies and visual selection as predictors of realistic drawing skill. *Psychology of Aesthetics, Creativity, and the Arts*, 6(2):124–136.

[104] Pascale, D. (2003). A Review of RGB Color Spaces . . . from xyYto RGB.

[105] Pawar, A. (2012). Review of CBVIR in WWW. *Advanced Research in Computer Engineering & Technology*, 1(3):47–50.

[106] Pérez-Carpinell, J., de Fez, M. D., Baldoví, R., and Soriano, J. C. (1998). Familiar objects and memory color. *Color Research & Application*, 23(6):416–427.

[107] Pianta, R. C., Longmaid, K., and Ferguson, J. E. (1999). Attachment-based classifications of children's family drawings: psychometric properties and relations with children's adjustment in kindergarten. *Journal of clinical child psychology*, 28(2):244–55.

[108] Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., and Brazier, J. E. (2011). Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Social science & medicine (1982)*, 72(10):1717–27.

[109] Pu, H. (2005). A comparative analysis of web image and textual queries. *Online Information Review*, 29(5):457–467.

[110] Pylyshyn, Z. W. (2002). Mental imagery: in search of a theory. *The Behavioral and brain sciences*, 25(2):157–182; discussion 182–237.

[111] Regier, T. and Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in cognitive sciences*, 13(10):439–46.

[112] Robertson, P. (1988). Visualizing color gamuts: a user interface for the effective use of perceptual color spaces in data displays. *IEEE Computer Graphics and Applications*, 8(5):50–64.

[113] Rui, Y., Huang, T. S., Ortega, M., and Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655.

[114] Schlak, T. M. (2010). *Image Retrieval as Information Seeking Behaviour? Self-Categorizations of User Motivaitons to Retrieve Images*. Phd thesis, University of Pittsburgh.

[115] Schwarz, M. W., Cowan, W. B., and Beatty, J. C. (1987). An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics*, 6(2):123–158.

[116] Seliger, H. H. (2002). Measurement of memory of color. *Color Research and Application*, 27(4):233–242.

[117] Shek, D. T. L. and Ma, C. M. S. (2011). Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *The Scientific World*, 11:42–76.

[118] Simons, D. (2007). Inattentional blindness.

[119] Siochi, A. C. and Ehrich, R. W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. *ACM Transactions on Information Systems*, 9(4):309–335.

[120] Smith, A. R. (1978). Color gamut transform pairs.

[121] Smith, A. R. and Lyons, E. R. (1996). HWBA More Intuitive Hue-Based Color Model. *Journal of Graphics Tools*, 1(1):3–17.

[122] Smith, J. and Chang, S.-F. (1997a). Querying by color regions using the VisualSEEk content-based visual query system. *Intelligent multimedia information retrieval*, 7(3):23–42.

[123] Smith, J. R. and Chang, S.-F. (1997b). Visually searching the Web for content. *IEEE Multimedia*, 4(3):12–20.

[124] Smith, W. J. (1988). Standardizing colors for computer screens. In *Human Factors and Ergonomics Society Annual Meeting*, pages 1381–1385.

[125] Sun, X., Wang, C., Sud, A., Xu, C., and Zhang, L. (2013). MagicBrush: image search by color sketch. In *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 475–476, New York, New York, USA. ACM Press.

[126] Tang, W. and Tu, X. (2012). *Modern Clinical Trial Analysis*. Springer Science & Business Media.

[127] Tollmar, K., Yeh, T., and Darrell, T. (2006). IDiexis: Mobile image-based search on world wide web-a picture is worth a thousand keywords. *Proc. of Mobisys*.

[128] Town, C. and Sinclair, D. (2000). Content based image retrieval using semantic visual categories. *AT&T Laboratories Cambridge Report*, pages 1–15.

[129] Trojano, L. and Conson, M. (2008). *Neuropsychology and Behavioral Neurology*, volume 88 of *Handbook of Clinical Neurology*. Elsevier.

[130] Tseng, K.-Y., Lin, Y.-L., Chen, Y.-H., and Hsu, W. H. (2012). Sketch-based image retrieval on mobile devices using compact hash bits. In *Proceedings of the 20th ACM international conference on Multimedia - MM '12*, pages 913–916. ACM Press.

[131] Tseng, L.-c. (2012). *Modeling Users Contextual Querying Behavior for Web Image Searching*. PhD thesis, Queensland University of Technology.

[132] Uchikawa, K. and Ikeda, M. (1986). Accuracy of memory for brightness of colored lights measured with successive comparison method. *Journal of the Optical Society of America. A, Optics and image science*, 3(1):34–9.

[133] van den Broek, E., Hendriks, M., Puts, M., and Vuurpijl, L. (2003). Modeling human color categorization: Color discrimination and color memory. *Proceedings of the 15th BNAIC*, 2003:59–68.

[134] van den Broek, E., Kisters, P. M. F., and Vuurpijl, L. G. (2004). Design guidelines for a Content-Based Image Retrieval Color-Selection Interface. In *Proceedings of the conference on Dutch directions in HCI,*, page 14.

[135] van den Broek, E., Vuurpijl, L., Kisters, P., and von Schmid, J. (2002). Content-Based Image Retrieval: Color-selection exploited. In *Third Dutch-Belgian Information Retrieval Workshop, DIR-2002,*, pages 38–47. Interdisciplinary Centre for Law and IT.

[136] Venters, C. C., Hartley, R. J., Cooper, M. D., and Hewitt, W. T. (2001). Query by Visual Example: Assessing the Usability of Content-Based Image Retrieval System User Interfaces. In Shum, H.-Y., Liao, M., and Chang, S.-F., editors, *Second IEEE Pacific Rim Conference on Multimedia,*, volume 2195 of *Lecture Notes in Computer Science*, pages 514–521, Berlin, Heidelberg. Springer Berlin Heidelberg.

[137] Venters, C. C., Hartley, R. J., and Hewitt, W. T. (2004). *Mind the Gap: Content-Based Image Retrieval and the User Interface*. IGI Global.

[138] Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A Survey on Pixel-Based Skin Color Detection Techniques. In *GRAPHICON-2003*, pages 85–92.

[139] Vuurpijl, L., Schomaker, L., and Van Den Broek, E. (2002). Vind(x): Using the user through cooperative annotation. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, (x):221–226.

[140] Wang, C., Li, Z., and Zhang, L. (2010). MindFinder: image search by interactive sketching and tagging. In *Proceedings of the 19th international conference on World wide web - WWW '10*, pages 1309–1312, New York, New York, USA. ACM Press.

[141] Wang, P., Zhang, D., Wang, J., Wu, Z., Hua, X.-S., and Li, S. (2012). Color filter for image search. *Proceedings of the 20th ACM international conference on Multimedia - MM '12*, page 1327.

[142] Wang, Y., Mei, T., Wang, J., Li, H., and Li, S. (2011). JIGSAW. In *Proceedings of the 19th ACM international conference on Multimedia - MM '11*, page 73, New York, New York, USA. ACM Press.

[143] Ware, C. (2004). *Information Visualization, 2nd Edition*. Elsevier Inc.

[144] Wells, E. and Tassinary, L. (1998). Standardized performance trajectory as a measure of usability. In *Proceedings Fourth Annual Symposium on Human Interaction with Complex Systems*, pages 226–233. IEEE Comput. Soc.

[145] Wells, E. F. (1994). *A Comparison of Interactive Color Specification Systems for Human-Computer Interfaces*. Phd thesis, Texas A&M University.

[146] Wijffelaars, M. (2008). Synthesis of color palettes. Master's thesis, Technische Universiteit Eindhoven.

[147] Wolfe, J. M. (1998). Visual memory: what do you know about what you saw? *Current biology : CB*, 8(9):R303–4.

[148] Wolfe, J. M. (1999). Inattentional amnesia. In Coltheart, V., editor, *Fleeting Memories: Cognition of Brief Visual Stimuli*, chapter 4, pages 71–94. MIT Press.

[149] Wu, Y. and Takatsuka, M. (2005). Three Dimensional Colour Pickers. In *Asia-Pacific Symposium on Information Visualisation (APVIS),*, pages 107–114. Australian Computer Society, Inc.

[150] Wyszecki, G. and Stiles, W. S. (1982). *Color science: concepts and methods, quantitative data and formulas*. Wiley.

[151] Xue, S., Tan, M., McNamara, A., Dorsey, J., and Rushmeier, H. (2014). Exploring the use of memory colors for image enhancement. In *Human Vision and Electronic Imaging XIX*, volume 901411.

[152] Zhang, D. (2010). A Survey of User Interfaces in Content-based Image Search Engines on the Web. In *Proceedings of the Fourth Workshop on HCI and IR(HCIR)*, pages 129–132.

[153] Zhang, M., Atkinson, S., Alechina, N., and Qiu, G. (2013). *A Is for Art My Drawings, Your Paintings*, volume 8158 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[154] Zhang, M., Atkinson, S., Qiu, G., and Alechina, N. (2014a). Can People Finger-draw Color-sketches from Memory for Painting Search on Mobile Phone? In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia - MoMM '14*, pages 115–118, New York, New York, USA. ACM Press.

[155] Zhang, M., Qiu, G., Alechina, N., and Atkinson, S. (2014b). A User Study on Color Palettes for Drawing Query on the Touchscreen Phone. In *MHCI'14*.

[156] Zhang, M., Qiu, G., Alechina, N., and Atkinson, S. (2015). A Comparison of Five HSV Color Selection Interfaces for Mobile Painting Search. In *15th International Conference on Human-Computer Interaction (INTERACT)*, Bamberg.