Name: Adrian Moise
Degree: Doctor of Philosophy
Title of Thesis: Designing Better User Interfaces for Radiology Interpretation
Examiner Committee:

Chair: Dr. Torsten Möller
Assistant Professor
Dr. M. Stella Atkins
Senior Supervisor
Professor
Dr. Arthur Kirkpatrick
Supervisor
Assistant Professor
Dr. Linda R. Bartram
Supervisor
Adjunct Professor, Department of Mathematics
Research Scientist, Colingo Networks Inc.
Dr. Elizabeth Krupinski
External Examiner
Research Professor, Department of Radiology and Psychology
University of Arizona
Dr. Christine MacKenzie
S.F.U. Examiner
Professor of Kinesiology
Date Approved: August 14, 2003
PARTIAL COPYRIGHT LICENCE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay:

Designing better user interfaces for radiology interpretation

Author: ______________________________
(Signature)

August 28, 2003
(Date Signed)
Abstract

Since the 1980s, radiologists have started to interpret digital radiographs using modern computer systems, a process known as softcopy reading. For softcopy reading, Hanging Protocols are used to automatically arrange images for interpretation upon opening a case, thus minimizing the need for physicians to manipulate images. We have developed a strategy, called HP++, which extends current hanging protocols with support for ‘scenario-based’ interpretation, matching the radiologist’s workflow and ensuring a chronological presentation of information.

We hypothesized that HP++ significantly reduces off-image eye fixations, the interpretation time, and the frequency and complexity of user input. We validated our hypothesis with inexpensive usability studies based on an abstraction of the radiologist’s task, transferred to novice subjects. For a radiology look-alike task, we compared the performance of 20 graduate students using our HP++ based interaction technique with their performance using a conventional interaction technique. We observed a 15% reduction in the average interpretation time using the staged approach, with one third fewer interpretation errors, two thirds fewer mouse clicks, and over 65% less eye gaze over the workstation controls. User satisfaction with the staged interface was significantly higher than with the traditional interface. Preliminary external validation of these results with physician subjects indicate our usability results transfer to radiology softcopy reading. We conclude that designing radiology workstations with support for HP++ can improve the performance of workstation users.
Dedication

In loving memory of my father,
who taught me to strive for enlightenment,
and of my grandmother,
who selflessly dedicated her life to our happiness.

Perfer et obdura; dolor hic tibi proderit olim.

-Ovid
Acknowledgments

My first thanks go to my wife, Florentina, for her loving support and encouragement throughout this long journey. She was my compass when my focus went astray. I am forever indebted to Florentina for giving birth to our bundle of joy, Isabela, whose love and affection provided warmth and nourishment to my heart.

Dr. Stella Atkins has been my mentor and my friend. Her breadth of interests and her perspective on successful research have been greatly appreciated. She gracefully motivated and supported my research, which draws from so many disciplines of which I knew initially so little. Her assistance in gaining access to key medical professionals was paramount to the success of this work.

I am very grateful to Dr. Arthur Kirkpatrick and Dr. Linda Bartram for their insightful comments and advice on methodology which helped greatly in improving the quality of this thesis. I am indebted to Dr. Alan Rowberg for bringing a visionary physician’s expertise to our line of work. His comments on the experimental setup provided invaluable support for transferring our results to the radiological domain.

I thank the other members of my Examining Committee, Dr. Elizabeth Krupinski and Dr. Christine MacKenzie for agreeing to examine a thesis which cuts across computer science, psychology and radiology.
I thank Dr. Vince Di Lollo, from the Department of Psychology at the University of British Columbia. His expertise in vision and visual attention was invaluable for guiding my search for appropriate stimuli for my experiment.

Over the years the members of the SFU Medical Computing Lab have provided a delightfully friendly and rich working environment. Many thanks to Anne Fouron for her assistance in running the pilot experiment, and to Ben Law for sharing his expertise on eye trackers.

I thank the people at A.L.I. Technologies for the opportunity to learn from others and share ideas. I am especially grateful to Marcel Sutanto, my industrial supervisor, a role model with great people skills and strong technical knowledge.

Support for this research was funded by Industrial Research Scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and BC Science Council, and by various graduate awards from Simon Fraser University.
Table of Contents

Approval ........................................................................................................................................ ii
Abstract .......................................................................................................................................... iii
Dedication ......................................................................................................................................... iv
Acknowledgments .......................................................................................................................... v
Table of Contents .......................................................................................................................... vii
List of Tables ....................................................................................................................................... xii
List of Figures ..................................................................................................................................... xiv
Abbreviations and Acronyms ........................................................................................................... xviii

1 Introduction ................................................................................................................................. 1
  1.1 Goals and motivation ............................................................................................................... 1
  1.2 Thesis outline .......................................................................................................................... 3

2 Radiological Softcopy Interpretation – Overview .................................................................. 5
  2.1 Picture Archiving and Communication Systems (PACS) .................................................. 5
    2.1.1 Evolution of PACS .......................................................................................................... 5
    2.1.2 Large scale PACS implementation ................................................................................ 8
    2.1.3 Advantages of the film based practice ......................................................................... 8
    2.1.4 Potential disadvantages of the use of PACS ................................................................. 9
    2.1.5 Advantages of the use of PACS .................................................................................... 10
  2.2 Industry Standards .................................................................................................................. 16
    2.2.1 Digital Imaging and Communication in Medicine (DICOM) ................................ 17
    2.2.2 Health Level Seven (HL7) ............................................................................................ 18
    2.2.3 Integrating the Healthcare Enterprise (IHE) .............................................................. 18
    2.2.4 Other standards ............................................................................................................. 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>Radiologic image acquisition systems</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>The Era of Huge Studies</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>Radiology workstation Design</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Hardcopy interpretation workflow</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Viewing mode</td>
<td>33</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Tile vs. stack</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>Radiological softcopy workflow: task decomposition</td>
<td>37</td>
</tr>
<tr>
<td>3.4</td>
<td>Radiology workstation tools and features</td>
<td>41</td>
</tr>
<tr>
<td>3.5</td>
<td>Workstation tools usage patterns</td>
<td>45</td>
</tr>
<tr>
<td>3.6</td>
<td>Classification of PACS workstations based on their role</td>
<td>46</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Primary diagnostic workstation</td>
<td>47</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Secondary review workstation</td>
<td>48</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Tertiary workstation</td>
<td>48</td>
</tr>
<tr>
<td>3.7</td>
<td>Design</td>
<td>49</td>
</tr>
<tr>
<td>3.8</td>
<td>Evolutionary stages of radiology workstation</td>
<td>50</td>
</tr>
<tr>
<td>3.9</td>
<td>Identifying the representative users</td>
<td>51</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Role and hospital department</td>
<td>52</td>
</tr>
<tr>
<td>3.9.2</td>
<td>Training level</td>
<td>53</td>
</tr>
<tr>
<td>3.9.3</td>
<td>Level of experience with radiology workstations</td>
<td>54</td>
</tr>
<tr>
<td>3.9.4</td>
<td>Organization</td>
<td>55</td>
</tr>
<tr>
<td>3.9.5</td>
<td>Area of expertise (specialty)</td>
<td>55</td>
</tr>
<tr>
<td>3.9.6</td>
<td>Body part</td>
<td>56</td>
</tr>
<tr>
<td>3.10</td>
<td>Case Study: Radiology workstation design for the ICU</td>
<td>58</td>
</tr>
<tr>
<td>3.10.1</td>
<td>Specific conditions in the ICU</td>
<td>58</td>
</tr>
<tr>
<td>3.10.2</td>
<td>ICU specific features for the PACS workstation</td>
<td>59</td>
</tr>
<tr>
<td>3.10.3</td>
<td>Reliable and fast -- new toolkit and user interface</td>
<td>60</td>
</tr>
<tr>
<td>3.10.4</td>
<td>Automation of most routine tasks</td>
<td>61</td>
</tr>
<tr>
<td>3.11</td>
<td>Workstation evaluation</td>
<td>61</td>
</tr>
<tr>
<td>3.12</td>
<td>Usability</td>
<td>62</td>
</tr>
<tr>
<td>3.13</td>
<td>Ergonomic user interfaces for the radiology workstation</td>
<td>63</td>
</tr>
<tr>
<td>3.14</td>
<td>Providing a consistent metaphor</td>
<td>63</td>
</tr>
<tr>
<td>3.15</td>
<td>Hanging Protocols</td>
<td>64</td>
</tr>
<tr>
<td>3.15.1</td>
<td>Problem statement</td>
<td>64</td>
</tr>
<tr>
<td>3.15.2</td>
<td>Hanging protocols: Definition</td>
<td>65</td>
</tr>
<tr>
<td>3.15.3</td>
<td>Related relevant work</td>
<td>65</td>
</tr>
<tr>
<td>3.15.4</td>
<td>Current limitations and challenges associated with HP</td>
<td>67</td>
</tr>
<tr>
<td>4</td>
<td>‘Historical evolution’ of our line of work</td>
<td>70</td>
</tr>
<tr>
<td>4.1</td>
<td>Hands-on industrial experience and direct interaction with the radiologists</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>Medical imaging conferences and trades shows</td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>First user study: radiology workstation design</td>
<td>71</td>
</tr>
<tr>
<td>4.3.1</td>
<td>User profiles</td>
<td>71</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Softcopy vs. hardcopy</td>
<td>73</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Hierarchical HP</td>
<td>76</td>
</tr>
</tbody>
</table>

*Designing Better User Interfaces for Radiology Workstations*
4.4 Softcopy interpretation: a literature survey .................................................. 81
4.5 Determining the requirements of diagnostic workstations .................... 81
4.6 Perception ...................................................................................................... 82
4.6.1 Understanding the visual search ............................................................... 83
4.6.2 Satisfaction of search ................................................................................ 85
4.6.3 Understanding the nature of expertise ....................................................... 87
4.6.4 Developing aids to image perception ......................................................... 88
4.7 Key components of the radiologist’s task ..................................................... 89
4.8 Usability engineering .................................................................................... 90
4.8.1 Classification of radiology workstation controls ..................................... 91
4.8.2 Case study: radiology workstation design for the ICU ............................ 91
4.9 DICOM WG11 and Supplement 60 on Hanging Protocols ....................... 92

5 Proposed solution: HP++ ............................................................................... 94
5.1 Study driven vs. scenario driven interpretation ........................................... 94
5.2 Scenario-based, workflow oriented hanging protocols with stages .......... 95
5.2.1 Potential advantages of the use of stages ................................................ 96
5.3 Handling heterogeneous data .................................................................... 97
5.4 Virtual desktops .......................................................................................... 98
5.5 Type abstraction hierarchy and approximate matching ............................ 99
5.6 Potential limitations for the use of HP++ .................................................... 102

6 Pilot experiment ............................................................................................. 104
6.1 A radiology look-alike task ......................................................................... 104
6.1.1 Motivation .............................................................................................. 104
6.1.2 Real-life radiology scenario .................................................................... 105
6.2 Purpose ......................................................................................................... 108
6.3 Hypothesis ................................................................................................... 109
6.4 Conjunctive visual search .......................................................................... 110
6.4.1 Target description .................................................................................. 111
6.5 Stimuli .......................................................................................................... 115
6.6 Procedure ..................................................................................................... 115
6.7 Independent variables ................................................................................ 122
6.8 Dependent measures .................................................................................. 122
6.8.1 ASL 504 Eye Tracker ............................................................................. 123
6.9 Pilot experiment results ............................................................................... 124
6.9.1 Response Time (RT) .............................................................................. 124
6.9.2 Mouse clicks (MC) ................................................................................. 125
6.9.3 Interpretation errors ................................................................................ 125
6.9.4 Eye gaze ................................................................................................ 126
6.9.5 NASA Taskload Index (TLX) ................................................................. 127
6.10 Subjective comments .................................................................................. 128
6.11 Discussion and lessons learned from the pilot ......................................... 129
# User study

## Experiment overview

- **7.2.1** Hypothesis
- **7.2.2** Task description
- **7.2.3** Procedure
- **7.2.4** The stimuli
- **7.2.5** Independent variables
- **7.2.6** Dependent measures

## Results

- **7.3.1** Response Time (RT)
- **7.3.2** Mouse clicks (MC)
- **7.3.3** Interpretation errors
- **7.3.4** Subjective ratings
- **7.3.5** Subjective comments
- **7.3.6** Discussion
- **7.3.7** Improving our interaction technique

## Feedback from radiologists on the user study

- **7.4.1** Medical Doctor
- **7.4.2** Radiologist
- **7.4.3** About the performance of our physicians

# Conclusions

- **8.1** Major contributions
- **8.1.1** Radiological interpretation: softcopy vs. hardcopy reading
- **8.1.2** Task analysis for radiological softcopy interpretation
- **8.1.3** Heads-up, streamlined radiology interpretation
- **8.1.4** Pioneers for inexpensive usability studies
- **8.1.5** HP++ and the automation of routine tasks
- **8.1.6** Radiologist-centred interaction technique
- **8.1.7** DICOM WG11 and Supplement 60 on Hanging Protocols

## Future work

## Summary

## Designing Better User Interfaces for Radiology Workstations
Appendix C: Examples on how various factors affect the trial complexity .......... 178
  Introduction: Shapes, targets, and distractors .................................................. 178
  Target enclosure and target contrast (relative to background) .......................... 178
  Target location .................................................................................................. 179
  Number of objects ............................................................................................ 179
  Registration ...................................................................................................... 180

Appendix D: Training script for the user study .................................................... 182

Appendix E: System Usability Scale (SUS) .......................................................... 185
  Using the SUS scale ......................................................................................... 186
  Scoring the SUS scale ..................................................................................... 186
  Notes on construction of the SU scale ............................................................. 186

Appendix F: Second usability questionnaire (Q19) ............................................ 188

Appendix G: Consent forms ................................................................................ 190
  First user study, November 1999 ................................................................. 190
  Pilot experiment ............................................................................................... 191
  User study ....................................................................................................... 192

Appendix H: Approvals from the Office of Research Ethics ............................... 194

References .......................................................................................................... 196
List of Tables

Table 1: A comparison between expert and novice users .................................................. 55
Table 2: Folder manager comparison: hardcopy vs. softcopy ........................................ 66
Table 3: Profiles of the radiologists involved in our study ............................................. 72
Table 4: Drawbacks of the hardcopy interpretation mode ............................................. 73
Table 5: Drawbacks of the softcopy interpretation mode ............................................. 74
Table 6: Most used features for the radiology workstation as of December 1999 ........ 75
Table 7: Most useful new features required for the radiology workstation as of December 1999 ................................................................. 76
Table 8: Hardcopy vs. softcopy – quality of diagnosis and image organization time ........................................................................................................ 77
Table 9: Estimated number of HP and percentage of the studies covered ................. 79
Table 10: Effect of display devices: general vs. specialized workstation; is frame effect disturbing? preferred number of monitors, and importance of spatial resolution .............................................................................................. 80
Table 11: Mean percentages of image areas receiving fixation clusters during search for CRT monitor vs. film viewing of skeletal radiographs. From (Krupinski and Lund 1997). ................................................................. 84
Table 12: Median dwell times (ms) for decisions made during search of skeletal images for trauma on film vs. a CRT monitor (6 readers). Only true-negatives differ significantly. From (Krupinski and Lund 1997) .......... 85
Table 13: Four stages for interpretation of a CR chest exam, when a prior study is available for comparison ................................................................. 108
Table 14: In Stages, each control corresponded to a pair of images to be displayed on screen ........................................................................................................ 117
Table 15: In FUI, each control corresponded to a single image ..................................... 119
Table 16: Counterbalancing for the interaction technique and the two sets of 15 stimuli each ................................................................. 134

Table 17: The three levels of trial complexity were generated by varying the target conspicuity, the configuration similarity, and the target location........ 140

Table 18: Rating system for configuration similarity.................................................. 141

Table 19: Paired Samples Statistics for Mouse Clicks ................................................. 147

Table 20: Interpretation errors by type ................................................................. 148

Table 21: Error distribution for male subjects. Two subjects made no interpretation errors ................................................................. 148

Table 22: Error distribution for the 10 female subjects. Two subjects made no interpretation errors ................................................................. 149

Table 23: The performance for the 2 subjects that preferred FUI .......................... 151

Table 24: Productivity, Quality of diagnosis & Time spent arranging the study (%) ......................................................................................... 168

Table 25: Present Drawbacks, Most Used Features and Proposed New Features ........ 169

Table 26: Proposed New Features vs. Importance .................................................. 170

Table 27: Templates ......................................................................................... 171

Table 28: TLX rating scale definitions ................................................................. 177

Table 29: System Usability Scale Questionnaire....................................................... 187
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Film alternator</td>
<td>2</td>
</tr>
<tr>
<td>Figure 2</td>
<td>First MPR. ©2003 Cedara Software Corp.</td>
<td>12</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Second MPR, using the same grey levels. ©2003 Cedara Software Corp.</td>
<td>12</td>
</tr>
<tr>
<td>Figure 4</td>
<td>CT chest, first volume rendering. ©2003 Cedara Software Corp.</td>
<td>13</td>
</tr>
<tr>
<td>Figure 5</td>
<td>CT chest, second volume rendering. ©2003 Cedara Software Corp.</td>
<td>13</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Window width and Level Mapping (W/L)</td>
<td>14</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Computed tomography slice with ‘lung’ display parameters (W=1500, L=-550)</td>
<td>15</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Computed tomography slice with ‘liver’ display parameters (W=270, L=125)</td>
<td>15</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Standard PA image</td>
<td>16</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Soft tissue image</td>
<td>16</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Bone image</td>
<td>16</td>
</tr>
<tr>
<td>Figure 12</td>
<td>A modern FCR 5000 CR Reader from Fuji.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 13</td>
<td>AXIOM Aristos FX DR system (courtesy: Siemens)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 14</td>
<td>MAGNETOM Trio 3 Tesla whole body MR system for highest performance</td>
<td>22</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Axial view of the activation map (Courtesy: Dr. Jeff Orchard)</td>
<td>23</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Sagittal view of the activation map (Courtesy: Dr. Jeff Orchard)</td>
<td>23</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Coronal view of the activation map (Courtesy: Dr. Jeff Orchard)</td>
<td>23</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Original ‘Siretom’ dedicated head CT scanner, circa 1974 (Courtesy: Siemens)</td>
<td>24</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Award-winning CT scanner (Courtesy: Siemens)</td>
<td>24</td>
</tr>
</tbody>
</table>
Figure 20: Original axial CT image (128 x 128) from the Siretom CT scanner (1975). (Courtesy: Siemens) ................................................................. 24

Figure 21: Axial CT image of a normal brain (512 x 512) (Courtesy: Siemens) .......... 24

Figure 22: A nuclear medicine study for cardiology (Courtesy: Siemens) ............... 25

Figure 23: Lower gastro-intestinal barium enema (Courtesy: Siemens) ................. 26

Figure 24: SONOLINE G60 STM ultrasound platform (Courtesy: Siemens) .......... 27

Figure 25: US Doppler used to visualize an ICA Stenosis (Courtesy: Siemens) ........ 27

Figure 26: Sequences with non-parallel axis ................................................................................................. 35

Figure 27: Only the top-most slice from the stack is labeled by the user...................... 57

Figure 28: The second slice from the stack is automatically labeled............................. 57

Figure 29: Another slice labeled automatically ....................................................................................... 57

Figure 30: In the right-most image, the arrow points to the location selected by the user. The software automatically calculates the corresponding location in two axial series................................................................. 57

Figure 31: Hierarchical HP organization ......................................................................................... 78

Figure 32: Current chest CR study of a patient shows suture marks............................ 106

Figure 33: Prior study of the same patient, showing no suture marks......................... 106

Figure 34: Targets ................................................................................................................................. 113

Figure 35: The same target is incompletely presented on two different images. Registration is required for solving the 'wild card' into a target. The wild card instantiates in a disc with the proper divider orientation. ......................... 114

Figure 36: The wild card can instantiate in a disc with incorrect divider orientation, thus representing a distractor................................................................. 114

Figure 37: Distractor, because the two wild cards do not transform into a split disc ........................................................................................................ 114

Figure 38: Study 1, first image ......................................................................................................... 115

Figure 39: Study 1, second image ................................................................................................. 115

Figure 40: Study 2, first image ......................................................................................................... 115

Figure 41: Study 2, second image ................................................................................................. 115

Figure 42: The two images from study 1 are shown from a trial using Stages. The target is located in the cross......................................................................................... 118

Figure 43: The two images from study two are shown from a trial with FUI. The target is in the cross................................................................................................. 119

Figure 44: An empty screen shows no stimuli. The only enabled button is 'Start Trial' .................................................................................................................... 120

Designing Better User Interfaces for Radiology Workstations xv
Figure 45: This layout corresponds to Stage 3 of a trial, showing the PA images from study 1 and 2. ................................................................. 121

Figure 46: ASL 504 remote eye-tracker .......................................................... 123

Figure 47: Calibration grid composed of 9 reference points uniformly distributed on the screen ................................................................. 124

Figure 48: Response time for each of the four subjects, under both IT conditions. Error bars represent one standard deviation. ......................... 125

Figure 49: Number of mouse clicks for each of the four subjects, under both IT conditions. Error bars represent one standard deviation. ......................... 125

Figure 50: Average distribution of eye-gaze .................................................... 126

Figure 51: Distribution of eye gaze in seconds of the main elements of the screen layout ................................................................................. 127

Figure 52: Distribution of the eye-gaze for the 4 subjects, under the two IT conditions ................................................................................. 127

Figure 53: TLX ratings for the four subjects, under both IT conditions ............. 128

Figure 54: An empty screen shows no stimuli. The only enabled button is ‘Start Trial’ ......................................................................................... 135

Figure 55: This layout corresponds to Stage 3 of a medium complexity trial, and shows the PA images from study one and two. Each image has 4 distractors and 3 wild cards .................................................... 137

Figure 56: An empty screen displayed after the completion of a trial. The radio buttons used to collect subjective ratings are the only controls enabled. ..... 138

Figure 57: Study 1, PA view ............................................................................ 141
Figure 58: Study 1, LAT view ....................................................................... 141
Figure 59: Study 1, PA view ....................................................................... 142
Figure 60: Study 1, LAT view ....................................................................... 142
Figure 61: Study 1, PA view ....................................................................... 142
Figure 62: Study 1, LAT view ....................................................................... 142

Figure 63: Average response time, calculated on both time and ln(t) coordinates. Error bars represent one standard deviation. ......................... 144

Figure 64: RT plotted against trial outcome .................................................... 145
Figure 65: RT plotted against trial complexity ................................................ 145
Figure 66: IT* METHOD had a significant effect on RT (p<0.001). ...................... 146

Figure 67: Number of errors per outcome condition ...................................... 149

Figure 68: Potential targets formations containing a wild card, and thus requiring the inspection of the second image ................................. 178
Figure 69: Distractors (non-targets, requiring NO inspection of the second image)...... 178
Figure 70: The eight shapes, with frames ................................................................. 179
Figure 71: The eight shapes, without frames .............................................................. 179
Figure 72: Target located in the center (open area) .................................................. 179
Figure 73: Target in bottom rectangle ...................................................................... 179
Figure 74: Max number of shapes is 8 (plus open area). The rule is 1-3 shapes requiring registration, plus another 4-6 requiring no registration ................. 180
Figure 75: Min number of shapes is 5. Only 1 shape requires registration, plus another 4 not requiring registration ......................................................... 180
Figure 76: This is an extreme example of difficult registration. Compared to the next image, objects shifted a lot, and the big circle disappeared, and two shapes lost their frames ............................................................. 180
Figure 77: The second image, used to show a difficult registration task in conjunction with the previous image .......................................................... 180
Figure 78: Easy registration with the next image ...................................................... 181
Figure 79: Easy registration with the previous image .............................................. 181
Abbreviations and Acronyms

- ACR: American College of Radiology
- ANOVA: Analysis of Variance
- CR: Computed Radiography
- CT: Computed Tomography
- DDP: Default Display Protocols
- DICOM: Digital Imaging and Communication in Medicine
- DR: Digital Radiography
- ECT: Emission computed tomography
- HCI: Human-Computer Interaction
- HIS: Hospital Information System
- HL7: Health Level Seven
- HP: Hanging Protocol
- IHE: Integrating the Healthcare Enterprise
• LOINC: Logical Observation Identifier Names and Codes

• MRI: Magnetic Resonance Imaging

• NM: Nuclear Medicine

• PACS: Picture Archiving and Communication Systems

• PET: Positron Emission Tomography

• RF: Radio-fluoroscopy

• RIS: Radiology Information System

• SPECT: Single photon emission computed tomography

• SNOMED: Systematized Nomenclature of Medicine

• UI: User Interface

• US: Ultrasonography

• W/L: window width and level

More abbreviations are available through the RAINFO LEXICON at http://www.rainfo.com/RAINFO%20LEXICON.htm.
1 Introduction

1.1 Goals and motivation

The radiologist’s primary task is to accurately identify various anatomic structures and pathological findings on medical images. For many years, the radiological images required for diagnosis and treatment planning were viewed on transparency film illuminated by a lightbox, called hardcopy reading. Some of the most time consuming tasks from the hardcopy interpretation workflow include:

- Retrieval of prior examinations from the film archive and associated reports from the radiology information system, which is generally done by the film clerk.
- Hanging the images for interpretation, which is generally done by the technologist.
- Spelling and formatting of reports, which is generally done by the transcriptionist.

Many new problems have arisen as radiologists progressed from reading images presented on film to images presented on high-resolution displays using modern computer systems, called softcopy reading. Although the digital medium has many advantages, the radiologist’s job became cluttered with new tasks related to retrieval, navigation and manipulation of a large amount of information. The radiologist may be forced to spend a greater percentage of time manipulating, rather than reading, the images. Since in many institutions radiologists are remunerated solely on the number of studies on which they report, any technology or workflow alteration which requires extra manipulation time for the radiologist is not acceptable. Hence, one can understand why...
digital diagnostic interpretation “is making life easier for everybody but the radiologist,” as one radiologist complained in 1999.

A successful transition from film-based reading to softcopy reading is conditioned, to a large extent, by replacing many of the functions done manually with automated equivalents. Our goal is to develop a strategy that models the standard radiological workflow, significantly reducing off-image eye fixations, the frequency and complexity of user input, as well as optimizing many other highly repetitive tasks involved in image navigation and manipulation.

One such routine task is setting up the presentation of radiological data for interpretation. In a film-based environment, prior to examination by a radiologist, a technician manually hangs the filmed examinations on the lightbox in a known examination-specific default (Blume, Bergstrom et al. 1998). For viewing of large numbers of films, an ‘alternator’ can be used to display up to 8 films at a time, as shown in Figure 1. Alternators are well suited for reviewing large numbers of films for complicated cases.

Figure 1: Film alternator

For softcopy interpretation the technician’s work is replaced by hanging protocols that are meant to partially reduce or completely eliminate the time spent by radiologists rearranging images on computer monitors (Strickland and Allison 1995). By hanging protocols (HPs) we mean a set of rules defining how images are arranged on one or more computer monitors upon opening a radiology case. The automatic image presentation
streamlines the radiologist's interpretation task. HPs usually include information regarding screen layout, image placement and image presentation.

The concept of hanging protocols extends the idea of optimizing the softcopy reading task rather than merely emulating the use of a film alternator, thus providing the potential for faster reading and improved diagnostic confidence. Other potential advantages include the development of guidelines for softcopy reading of radiographic studies and practices for optimal readout of specific cases. The standardization of the description of these protocols would allow institutions to share protocols for comparison and evaluation. However, current HPs provide just the initial presentation of images and do not handle other clinical information normally reviewed by radiologists, such as prior radiology reports (Moise and Atkins 2002c).

In order to address these limitations, we developed a method called HP++ that extends current hanging protocols with support for 'scenario-based' interpretation, which matches the radiologist's workflow and ensures a chronological presentation of information. Using task decomposition we identified the sequence of interpretation steps associated with softcopy reading, where each step requires unique data presentation (Aberle, Dionisio et al. 1996; Dionisio, Cardenas et al. 1996; Bui, Aberle et al. 1998; Dayhoff and Kuzmak 1998; Bui, McNitt-Gray et al. 2001; Siegel and Reiner 2001; Ratib, Allada et al. 2002). We introduced the concept of 'stages' to describe what information is required, how this information is presented, and what tools are most likely to be used at each interpretation step. Correlated with approximate query sequence matching, HP++ can be customized to suit various medical specialties, personal preferences, and patient diagnoses. Integrating HP++ concepts in the user-workstation interaction can increase radiologists' productivity by reducing disruptions of their analysis, particularly disruptions of visual search.

1.2 Thesis outline

Chapter 1 presents some challenges associated with softcopy reading, thus providing a motivation for this thesis. Chapter 2 reviews historical milestones and current research in radiological softcopy interpretation, and also introduces the reader to image acquisition systems and relevant medical standards. In Chapter 3 we focus on the
radiology workstation: classification, representative users, and new design trends. Chapter 4 provides a historical perspective of our line of work.

Chapter 5 introduces HP++, our adaptive extension to HPs that handles stages, heterogeneous data, virtual desktops and approximate matching. The following two chapters focus on providing evidence for the benefits of stages and scenario-driven interpretation. Our pilot experiment and main user study are presented in Chapter 6 and Chapter 7, respectively. The thesis concludes with the listing of our major contributions, the summary and some future directions for this work.
2 Radiological Softcopy Interpretation – Overview

2.1 Picture Archiving and Communication Systems (PACS)

For the first 100 years since the discovery of x-rays, film was used almost exclusively for radiological interpretation. Medical imaging technology has been largely concerned with conventional film-screen x-ray imaging. Lately we have been witnessing the development and commercialization of several new imaging technologies. As available technology continues to evolve, many healthcare organizations are looking to PACS to move images beyond the walls of the radiology department.

PACS represents a filmless replacement of the workflow in a medical imaging enterprise. PACS is responsible for the management of radiological data, which primarily consists of images. PACS functionality includes image acquisition from modalities, transfer to image servers, storage, retrieval and prefetching using a hierarchic archive, and display on diagnostic workstations. Newer generations of PACS are tightly integrated with the hospital information system (HIS) and the radiology information system (RIS) via patient demographic information and diagnosis reports, and integrated with the imaging modalities via patient worklists.

2.1.1 Evolution of PACS

The PACS acronym was proposed by Dr. Judith Prewitt and used in the first PACS Conference (Duerinck, 1982). The list of alternate names includes ‘image...
management and communication system' and 'integrated diagnostic system' (Dwyer 2000).

The first attempts to create a digital solution for the communication, archiving and display of medical images started in early 1980s, but the progress was slow due to lack of appropriate and affordable technical solutions, cumbersome workstations, lack of interfaces to other clinical information systems and immaturity in understanding the digital workflow (Mendenhall, Dewey et al. 2001). The technological limitations started to improve considerably a few years later, but the solutions offered were mostly incompatible with the logistics of medical practice.

Most of the initial systems were low-tech solutions developed in university departments, the first so-called mini-PACS. Mini-PACS are solutions for modalities that have special functions or requirements not supported by a general purpose PACS. For example, ultrasonography and nuclear medicine lend themselves to this scenario because these modalities generate data that may require special resources, such as colour displays and/or video playback. However, these modalities are also more amenable to going filmless, because the referring physicians rely on the report rather than primary viewing.

The implementation of the first mini-PACS started in 1978 at the University of Kansas, interfacing 6 imaging modalities. The workstations used INTEL 8086/8087 processors running MP/M-86 operating system. The maximum network throughput was 850kb/s, which was too slow to satisfy the requirements of PACS (Dwyer 2000). Based on proprietary implementations, each mini-PACS functioned independently, lacking cooperation and connectivity with other hospital systems (Carrino, Unkel et al. 1998).

UNIX was the operating system used for the first generation of PACS, because UNIX was considered the main platform able to deliver enough computational resources. However, with the introduction in 1994 of Microsoft Windows NT and Intel's Pentium II processor, a new era arrived for diagnostic-quality workstations based on cheaper PCs. It became possible to distribute images more quickly, to more users, and at higher resolutions. All these advances created the context for the development of the second-generation PACS characterized by modular architecture, progressive implementation,

The Army, with much better financial resources, had several research projects that helped spread the PACS concept. Examples of such research projects include the teleradiology project from 1983, the Defense Imaging Network-Picture Archiving and Communications Systems (DIN-PACS) from 1985, and the Medical Diagnostic Imaging Support, also from 1985. Building on the success of the early research projects, DIN-PACS II from December 1997 allocated $7.3 million for the installation of a PACS network at Portsmouth Naval Medical Treatment Facility in Portsmouth, VA.

Early PACS implementations used a local cache on every workstation. In order to read images on a certain workstation, one had to copy first the images from the central server onto the local hard disk before displaying them on the workstation. Since it was not always easy to predict where a certain study would be read, the server had to broadcast images to all workstations. The major disadvantages with image broadcasting were increased server workload and network congestion. In order to prevent this situation, auto-routing algorithms were introduced to direct images only to the workstation where the study would be read.

When a PACS solution used modules from several vendors, transferring images between the imaging modalities and the main server was always an issue due to limitations in the vendor’s compliance to the medical imaging standards. For connectivity reasons, a slow syntax was often used to move images from a server belonging to vendor A to a workstation belonging to vendor B. Pulling images was much improved by using direct workstation access to the network location where images were stored.

Nowadays, technological advances such as 1GB/s network speeds and scalable storage solutions using Redundant Array of Independent Disks (RAIDS) and Network Attached Storage allowed for the implementation of ‘cache-less’ systems. Workstations can now retrieve and display the requested images in less than two seconds.
2.1.2 Large scale PACS implementation

Two surveys, one in 1993 and one in 1995 discussed the parameters for defining a large PACS (Bauman 1996a; Bauman 1996b). The first survey found 13 institutions with a PACS operational in November 1993, which met three criteria: in daily clinical use, three or more modalities, and images available inside and outside radiology. These systems had 5 to 17 image acquisition devices.

For the second survey, the definition of large PACS included an additional criterion, namely that a large PACS must handle a minimum of 20,000 studies annually. From the 23 large PACS identified as of February 1995, 12 were mentioned in the 1993 survey. Some of the interesting trends reported in this 1995 survey include:

- external access: one-third of the large PACS had more than 20 workstations for review outside of radiology,
- archiving: jukeboxes with optical disks were the standard means for long-term archiving,
- compression: all but one large PACS avoided lossy compression before interpretation.

2.1.3 Advantages of the film based practice

Most radiologists were trained to read images from films, and it is not rare for these physicians to resist changing their habits. In this section we will present some of the advantages of film interpretation which are most commonly used to justify the radiologists' reluctance to switch to softcopy reading. A description of the radiology workflow in a film environment is presented in Section 3.2.

In some institutions, the film interpretation was optimized through years of training, experience and incremental improvement of film-handling processes. Thus, the film workflow can provide a robust and complex manual system, with versatile exception handling. For example, if by accident an X-ray is printed on film with the wrong patient name, all it takes for the technician to fix this problem is to write on film with a permanent marker the correct patient name, and then insert the film the appropriate film jacket. In a PACS environment, correcting this problem requires (1) updating the PACS database to reflect the move of the digital image from one patient to another, which may
require system access rights at the PACS administrator level, and (2) updating the patient demographic information present in the header of the image itself.

During film interpretation, the radiologist’s visual tasks are interspersed with physical tasks, such as changing the film viewing distance or using a magnifier to examine the images. During softcopy reading, there is minimal physical movement, and increased mental workload for workstation manipulation.

Compared with the traditional film viewing on lightboxes, the monitors used for softcopy interpretation are less bright, have less special resolution, have less dynamic range, and have a limited viewing area. Lightboxes provide excellent luminance output, a high spatial resolution, and contrast range (Wong, Polunin et al. 1998). The light transmitted through a film at commercial film lightboxes has been measured at 520 fL, at a contrast ratio of 800:1 and can display an entire mammogram at a 4000 x 5000 pixels of resolution. The film lightboxes come in 2-4 banks of 4 films each, providing a significantly larger display area than computer monitors.

State-of-the-art CRT monitors can only display 2048 x 2560 pixels, with 8 or 10 bit/pixel contrast resolution, a luminance of up to 300 cd/m² and a contrast ratio of about 200:1. The images are usually acquired as 16 bit/pixel, but monitors can generally display 8 bit/pixel. Consequently, the radiologist has to adjust the brightness and contrast ratio to compensate for the reduced contrast resolution. The radiologist also has to zoom and pan, to compensate for the lower spatial resolution for images with a large matrix, such as mammograms. The radiologist is also required to navigate through the images on the workstation, to compensate for the reduced display area compared to film.

2.1.4 Potential disadvantages of the use of PACS

The start-up costs for the installation of PACS can delay its adoption. These costs include the hardware, laying the network, cost of training, and cost for software and maintenance.

Technical limitations, such as the flicker and fuzziness of cheap or defective monitors, can impose perceptual limitations on the radiologist. The radiologist also has to
deal with the mental load for interacting with the workstation for common operations such as zooming, panning, and adjusting the image contrast and brightness, which makes the usability of the diagnostic workstation a key factor in the acceptance of a PACS system.

It is no secret that technological capability does not guarantee proper operation (Horii, Farn et al. 2001; Horii, Redfern et al. 2001). One study indicated that PACS introduction slowed the technologist workflow, while the radiologist workflow was decreased or not affected (Redfern, Horii et al. 2000). Such disadvantages are not uncommon for a hybrid solution where PACS and film coexist, or where the digital solution is not fully implemented. The productivity will likely suffer because of a duplication of resources corresponding to the digital and film-based workflow.

2.1.5 Advantages of the use of PACS

Many imaging modalities are inherently digital, or are becoming digital, as we show in Section 2.3. PACS is expected to improve radiology department productivity and expedite consultation by radiologists. For computed tomography, the enhanced staff productivity due to the introduction of PACS was demonstrated by a 16% overall reduction in the radiologist time required for study interpretation, and a 45% reduction in the technologist time required for study acquisition (Reiner, Siegel et al. 1998b).

A description of the PACS-based radiology workflow is presented in Section 3.3. The main advantage of PACS is improvement in efficiency due to the management of digital data. The automation of routine tasks allows to display clinical history and radiological reports in association with corresponding images, as well as to retrieve and simultaneously display past and current images. PACS can be naturally integrated with the HIS/RIS system, thus eliminating manual data entry, which is error-prone. In order to correct demographic errors, facilitate quality control and insure proper routing of images, relay or modality quality control stations can be used to bridge the patient metadata information, managed by RIS/HIS, with the image data sets, managed by PACS (Carrino, Khorasani et al. 2000).
PACS is synonymous with greater reliability, speed, and convenience: images are available for viewing, teaching, or collaboration at multiple stations at the same time. For example, the US military model requires any image to be available on any computer within two seconds. PACS can be used to connect critical care areas, remote centres and clinics, radiologists’ homes and referring physicians’ offices, and to provide over-reads and subspecialty consultation via web technology (Strickland 1996).

PACS virtually eliminates lost or unavailable studies. For example, with the introduction of PACS at VA Baltimore, the lost film rate dropped from 23% to less than 1%, and unread film rate dropped from 8% to 0.3%. Other indirect benefits of PACS include a decreased rate of repeat and unnecessary studies, from 7.7% prior to PACS to about 3% afterwards at Hammersmith Hospital, and from 5% prior to PACS to about 0.8% afterwards at VA Baltimore Hospital (Strickland 1996). Furthermore, the courier costs and the unproductive physicians’ travel between centres can be reduced or eliminated. For example, after the introduction of PACS at VA Baltimore 98% of physicians saved 50-60 min per day (Erickson, Ryan et al. 1998).

There are also significant economical benefits associated with the use of PACS. For example, at the Baltimore VA Medical Center, the elimination of film, estimated at about $200,000 per year, resulted in a significant space saving by disposing of a 2,500 square foot file room (Erickson, Ryan et al. 1998). Film elimination made obsolete the clerks required to keep films organized, and also cut the cost for processor chemicals and waste disposal, heating and lighting of film file room, and film jackets repairs.

Other advantages of PACS include three-dimensional visualization and image enhancement, detailed below.

### 2.1.5.1 Three-dimensional visualization

Dr. Paul Eldridge, a consultant neurosurgeon at the Walton Center for Neurology and Neurosurgery in Liverpool, United Kingdom, made the following recent comment about the use of benefits of volume visualization for diagnostic interpretation:

Modern imaging techniques now acquire large amounts of data describing a volume. The traditional method of displaying data - a series of axial
slices printed onto a film examined on a viewing box - is now impractical. Volume rendering solves this problem by allowing visualization of anatomical and pathological relationships between structures (PRNewswire 2003).

Cross-sectional imaging techniques, such as magnetic resonance and computed tomography, produce series of two-dimensional images corresponding to a three-dimensional volume. Three-dimensional (3D) tools allow rapid and easy visualization of anatomical structures in ways that provide insights during radiological interpretation and surgical planning. The clinical benefits of 3D are becoming more recognized, and several modern workstations now offer tools for:

- multi-planar reformatting (MPR), as shown in Figure 2 and Figure 3. The original grey level can be textured in the cut surface or can be visualized as an MPR plane. This functionality can be used to correlate the location of the MPR planes shown in the application with the 3D visualization.
- tissue classification of multiple anatomical structures, as shown in Figure 4 and Figure 5. Using multiple tissue classifications an application can render with different colour and opacity anatomical structures which share the same densities.
- measurement of distance, angle, or volume,
- maximum intensity projection (MIP),
- virtual navigation through various anatomical structures, such as the colon.
The effect of PACS on the improved report turnaround time can be explained by interventions tailored for different departmental sections, such as electronic signature from the radiologist’s home, buddy system for signing reports, and structured reports (Seltzer, Kelly et al. 1997). The improved report turnaround time due to the introduction of PACS was reported by many studies:

- VA Baltimore: radiologists’ reporting times were reduced from several hours to less than 30 min, and the overall report time dropped from 26 hours to approximately 2h (Strickland 1996).
- Mayo Medical Center, Rochester: median 65 minutes with film reduced to 26 minutes with PACS (Matter, King et al. 1999).
- Harvard Medical School: 55% reduction in the overall report turnaround time, 80% reduction in the transcription time, and 68% reduction in the signature time (Reiner, Siegel et al. 2001).

Another important advantage of digital imaging is the ability to enhance selected images individually using image processing. This could lead to a higher detection probability and therefore a better diagnostic performance by radiologists. In this section we present two of the most commonly used algorithms for image enhancement: contrast enhancement and digital energy subtraction.

Medical images, whether directly acquired from digital imaging modalities or by scanning films, are typically recorded with 4096 intensities of gray. Thus, each pixel of
the image is represented by a 12-bit value. The acquired image may not always be satisfactory for radiologists’ interpretation: some images may be underexposed (too dark) while others may be overexposed (too bright). To improve readability, there is a need to adjust image brightness or contrast. Window and Level (W/L) is the solution to satisfy this need. It is a expansion of the contrast of the pixels within a given window range. The Window defines a range of values to map between black and white. The Level indicates a value to center the window on. The intensity of ‘uninterested’ pixels with values outside of the window are either mapped to black (fully underexposed) or white (fully overexposed). For example, in Figure 6 the window range is between 1024 and 3071. The pixel value of 2000 is within the window, while the value of 400 is outside of the window and therefore is mapped to ‘black’. The pixel value of 5000 is also outside of the window and is mapped to ‘white’. An example of the same medical image presented with two different W/L values is presented in Figure 7 and Figure 8, respectively.

Figure 6: Window width and Level Mapping (W/L)
Dual energy subtraction radiography exploits the energy dependence of x-ray attenuation by different tissues. The first practical subtraction technique used a single exposure detected by two receptor plates separated by a filter. A new subtraction technique uses a rapid sequence of two exposures, each at a different energy level. For example, for chest radiography the system automatically generates three posterior-anterior chest radiographs: a standard radiograph, an image of the soft tissues with the bones removed, and an image of the skeletal system together with any abnormal calcification that may be present, as shown in Figure 9, Figure 10, and Figure 11, respectively.
2.2 Industry Standards

An important rule in building a PACS system is to incorporate as many industry standards as possible that are consistent with the PACS architecture. Compliance to these standards will increase the portability and interoperability of the system, minimize the development of customized software, and reduce the effort for system maintenance.

Among American National Standards Institute accredited Standards Developing Organizations (SDO) operating in the healthcare arena, Health Level Seven (HL7, used for textual data) and Digital Imaging and Communication in Medicine (DICOM, used for medical images) are the most important (DeJarnette 1999). A SDO develops specifications for a messaging standard that enables disparate applications to exchange information.

Integrating the Healthcare Enterprise\(^1\) (IHE) is an initiative to stimulate the healthcare information and management systems industry to develop means for connecting and integrating clinical information systems, imaging systems and other information resources that, now, do not communicate with each other. IHE is spearheaded by the Healthcare Information and Management Systems Society and the Radiological Society of North America. The aim of IHE is to provide a highly visible showcase and forum for demonstrating the capabilities of a growing and evolving set of

\(^1\) http://www.rsna.org/IHE/index.shtml
standards. Its cost-effectiveness comes from eliminating duplication of efforts, cutting down on the likelihood that medical information will be lost or misplaced, and reducing the time and manpower needed to move data from one place to another.

### 2.2.1 Digital Imaging and Communication in Medicine (DICOM)

DICOM is a communications standard set by the American College of Radiology and the National Electronic Electrical Manufacturers Association. The purpose of DICOM is to provide uniformity and compatibility between medical imaging equipment vendors. Once an image is acquired at the host modality, it must be saved in a standard digital format before it can be sent over a computer network. In order to share patient demographics and other necessary image information, the vendors must comply with the DICOM standard.

The DICOM Standards Committee, formed in 1983, published in 1985 the first version of the standard. However, most of the promise of a standard method of communicating digital image information was not realized until the release of version 3.0 of the standard in 1993. The DICOM standard version 3.0 specified a network protocol, defined the operation of Service Classes beyond the simple transfer of data, and created a mechanism for uniquely identifying ‘Information Objects’ as they are acted upon across the network. DICOM defined information objects not only for images, but also for patients, studies, reports, and other data groupings. With these enhancements the standard was now ready to deliver on its promise to permit the transfer of medical images in a multi-vendor environment and to facilitate the development and expansion of PACS.

Working groups (WG) are formed by the DICOM Standards Committee to work on a particular classification of tasks. The 21 working groups perform the majority of work on extending and correcting the Standard. For the scope of this document, the most relevant activity is performed by WG11. WG11 is working on DICOM Supplement 60 on Hanging Protocols (as defined in Section 3.15), intended to allow physicians to conveniently define their preferred methods of presentation and interaction for different

---

2 The author has been a member of WG Eleven since December 2001.
types of viewing circumstances. After a physician has defined his/her preference once, the image sets will automatically be hung according to the user’s preferences on display systems of similar capability.

2.2.2 Health Level Seven (HL7)

HL7 is a standard for exchanging clinical and administrative data between various digital systems of the entire health care organization. ‘Level Seven’ refers to the highest level of the International Standards Organization’s communications model for Open Systems Interconnection - the application level. The application level addresses definition of the data to be exchanged, the timing of the interchange, and the communication of certain errors to the application. The seventh level supports such functions as security checks, participant identification, availability checks, exchange mechanism negotiations and, most importantly, data exchange structuring. Members of Health Level Seven are known collectively as the Working Group, which is organized into technical committees and special interest groups. The technical committees are directly responsible for the content of the Standards. Special interest groups serve as a test bed for exploring new areas that may need coverage in HL7’s published standards, such as Context Management.

Context management entails the coordination and synchronization of applications so they become mutually aware of the set of common things - known as the context - that frame and constrain the user’s interactions with applications. The HL7 Context Management Standard defines a protocol for securely ‘linking’ applications so that they ‘tune’ to the same context. Linked applications remain automatically synchronized even when a context subject changes, for example, due to the user selecting a different patient.

2.2.3 Integrating the Healthcare Enterprise (IHE)

Using subsets from DICOM and HL7, IHE defines the informational model that specifies how data must be created, managed, manipulated and exchanged to deliver radiology services to patients. The process-oriented IHE Technical Framework defines a set of actors and transactions, grouped in seven integration profiles: Scheduled Work Flow, Patient Information Reconciliation, Consistent Presentation of Images,

2.2.4 Other standards

HL7 categorizes messages so the system at the other end knows what to expect. For the information transferred to be useful, however, the data elements themselves must be standardized.

The purpose of the Logical Observation Identifier Names and Codes (LOINC3) database is to facilitate the exchange and pooling of results, such as blood haemoglobin and serum potassium, or vital signs for clinical care, outcomes management, and research.

The Systematized Nomenclature of Medicine (SNOMED4) is a comprehensive, multi-axial, controlled terminology created for the indexing of the entire medical record. SNOMED allows for consistent gathering of detailed clinical information, thus enabling providers of various specialties, researchers and even patients to share common understanding of health across sites of care and computer systems. JJ1017 is a Japanese extension of SNOMED, and resulted from the fact the Japanese image examination orders are much more specific than the American ones (Kimura, Kuranishi et al. 2002).

2.3 Radiologic image acquisition systems

About 70% of all radiological examinations, including skull, chest, breast, abdomen and bone, are acquired and stored on x-ray films (Reiner 2002). This process compresses a three-dimensional object into a two-dimensional image, and is called projection radiography. An x-ray film can be converted to a digital image using digitizers, such as TV camera digitizers, charged coupled device scanner digitizers or laser scanner digitizers.

3 http://www.loinc.org/
4 http://www.snomed.org/
Digital projection radiography can be acquired using computed radiography (CR), a technology that uses a laser-stimulated luminescence phosphor imaging plate as an x-ray detector. Eastman Kodak Company patented in 1975 an apparatus to store infrared-stimulable phosphors to store an image. Fuji Photo Film patented in 1980 the use of photo-stimulable phosphors to record an image. Fuji was also the first to commercialize a storage phosphor-based CR system in 1983 as the FCR 101. Figure 12 shows a modern CR reader from Fuji, equipped with a four cassette stacker and with a processing capacity of approximately 115 images per hour.

![Figure 12: A modern FCR 5000 CR Reader from Fuji.](image)

Digital radiography (DR) can capture images without going through the additional medium (imaging plate), as shown in Figure 13. By eliminating the film development step and the time to acquire the image from a CR reader, DR allows for a fast quality check. DR also provides a wide dynamic range. When combined with a radiology information system, these advantages can lead to a 30% reduction in total examination time over the traditional film practice (May, Deer et al. 2000). DR can also improve the patient response. For example, 67% of the patients involved in a study felt DR was superior to film (Dackiewicz, Bergsneider et al. 2000).
Magnetic Resonance Imaging (MRI) devices take images of objects through probing the magnetic moments of nuclei (usually protons), employing radio frequency radiation and strong magnetic fields. In 1975 Richard Ernst suggested using magnetic resonance imaging with phase and frequency encoding, and the Fourier Transform. His technique is the basis of current MRI techniques.

MRI started out as a tomographic imaging technique, which produced an image of the nuclear magnetic resonance signal in a thin slice through the human body. Using this technique in 1980, a single image could be acquired in approximately five minutes. By 1986, the imaging time was reduced to about five seconds, without sacrificing too much image quality. MRI has advanced beyond a tomographic imaging technique to a volume imaging technique. For example, Siemens's MAGNETOM Trio, shown in Figure 14, is the first generation of 3Tesla whole-body MR system. This Ultra High-Field system has an increased signal-to-noise ratio which is useful for brain imaging and orthopaedic imaging, by pushing the image resolution to achieve a 2 mm slice thicknesses.
In 1993 functional MRI (fMRI) was developed. This technique allows the mapping of the function of various regions of the human brain. The principle of fMRI imaging is to take a series of images of the brain in quick succession and to statistically analyze the images for differences among them. For example, in the blood-oxygen level dependent method, the fact that haemoglobin and deoxyhaemoglobin are magnetically different is exploited. Brain areas with more oxygenated blood flow are brighter on MRI images. Therefore, hyper-intense regions are thought to be correlated with brain activation. This has been exploited in fMRI where a series of baseline images are taken of the brain region of interest when the subject is at rest; then the subject performs a task and a second series is taken. Regions which show intensity fluctuations correlated with the task / resting sequence are presumed to have been activated by the task.

The result of the fMRI experiment is an activation map that highlights the regions of relevant brain activity. Figure 15, Figure 16, and Figure 17 show three orthogonal projections of an activation map overlaid on anatomical MR images of the brain. This activation map resulted from an experiment involving 72 snapshots acquired 3.5 seconds apart on a 1.5T Marconi scanner. The subject was observing a screen where a visual stimulus was displayed for 21 seconds, followed by 21 seconds of rest. This stimulus cycle was repeated 6 times in total.
Computed Tomography (CT) was invented in 1972 by the British engineer Godfrey Hounsfield from EMI Laboratories, England, and independently by the South African born physicist Allan Cormack of Tufts University, Massachusetts. Hounsfield was later awarded the Nobel Peace Prize and honored with Knighthood in England for his contributions to medicine and science. When the CT scanner was first introduced, it utilized a single x-ray beam as the energy source, and took 4.5 minutes to collect the necessary data to perform the single two-dimensional picture reconstruction. The helical CT scanner was introduced in late 1980s, based on three technological advances: the slip-ring gantry, improved detector efficiency, and greater x-ray cooling capacity. Nowadays one can get a head-to-toe CT scan in less than 15 minutes, with up to 4 images acquired in 1.6 seconds. Two moments from the evolution of the CT scanner are presented in Figure 18 and Figure 19. Figure 20 and Figure 21 show two moments from the evolution of CT image quality.
CT generates a series of images with pixel values reflecting Hounsfield units, where air has a value of -1000, water has the value 0, and bone has a larger positive value which is less than 3000 (Kuzmak, Reiner et al. 2000). The same CT slice from an abdominal scan is shown with two presentation parameters, corresponding to lung (Figure 7) and respectively to liver (Figure 8).

Nuclear medicine imaging documents organ function and structure, in contrast to diagnostic radiology, which is based upon anatomy. It is a way to gather medical information that may otherwise be unavailable, require surgery, or necessitate more expensive diagnostic tests. Nuclear medicine uses very small amounts of radioactive materials or radiopharmaceuticals to diagnose and treat disease. Radiopharmaceuticals are substances that are attracted to specific organs, bones, or tissues. The radiopharmaceuticals used in nuclear medicine emit gamma rays that can be detected.
externally by special types of cameras. These cameras work in conjunction with computers used to form images that provide data and information about the area of body being imaged, as shown in Figure 22. The amount of radiation from a nuclear medicine procedure is comparable to that received during a diagnostic x-ray.

Figure 22: A nuclear medicine study for cardiology (Courtesy: Siemens)

Positron Emission Tomography (PET) is currently the most effective way to check for cancer recurrences. Studies demonstrate that PET offers significant advantages over other forms of imaging such as CT or MRI scans in diagnosing disease. PET images demonstrate the chemistry of organs and other tissues such as tumours. A radiopharmaceutical which includes both sugar (glucose) and a radionuclide (a radioactive element) that gives off signals, is injected into the patient and its emissions are measured by a PET scanner. A PET scanner consists of an array of detectors that surround the patient. Using the gamma ray signals given off by the injected radionuclide, PET measures the amount of metabolic activity at a site in the body and a computer reassembles the signals into images. Cancer cells have higher metabolic rates than normal cells, and show up as denser areas on a PET scan. PET is useful in diagnosing
certain cardiovascular and neurological diseases because it highlights areas with increased, diminished or no metabolic activity, thereby pinpointing problems.

Single Photon Emission Computed Tomography (SPECT) is used to visualize functional information about a patient's specific organ or body system. Internal radiation is administered by means of a pharmaceutical which is labelled with a radioactive isotope. The radioactive isotope decays, resulting in the emission of gamma rays. These gamma rays provide information of what's happening inside the patient's body. SPECT uses gamma ray emissions as the source of radiation, rather than X-rays used in conventional Computed Tomography.

Fluoroscopy is a technique for obtaining "live" X-ray images of a living patient. The Radiologist uses a switch to control an X-Ray beam that is transmitted through the patient. The X-rays then strike a fluorescent plate that is coupled to an "image intensifier" that is (in turn) coupled to a television camera. The Radiologist can then watch the images "live" on a TV monitor. Fluoroscopy is often used to observe the digestive tract, as shown in Figure 23, or to observe the action of instruments being used either to diagnose or to treat the patient.

Figure 23: Lower gastro-intestinal barium enema (Courtesy: Siemens)
Ultrasonography (US) attempts to reconstruct a cross-sectional view of the patient by way of detecting the amplitudes of acoustical reflections that occur at the interface of tissues having different acoustical properties. Most US units, such as the one shown in Figure 24, have three sampling modes: the survey mode, in which the data stored in memory are continually updated and displayed, the static mode, in which only the maximum values during the scanning session are stored, and an averaging mode, in which data of all scans for a particular scan location are stored and displayed.

US using the Doppler principle can detect the movement of blood inside vessels, away or toward the scanning plane. Different colours can be used to code blood flow direction and speed with respect to stationary anatomical structures, as shown in Figure 25.

Figure 24: SONOLINE G60 S™ ultrasound platform (Courtesy: Siemens)

Figure 25: US Doppler used to visualize an ICA Stenosis (Courtesy: Siemens)
2.4 The Era of Huge Studies

New technologies are pushing the size of the data acquired. Horii et al. noted that the number of images obtained per examination in US increased with the use of PACS, partly because of changes in acquisition protocols (Horii, Farn et al. 2001). After the introduction of PACS at Baltimore VA Medical Center, the number of examinations per inpatient day increased by 82%, the number of examinations per visit increased by 21%, while the average length of stay decreased by 22% (Reiner, Siegel et al. 1998c). Technological advances allowed for a fast CT scanner to generate in axial mode up to 4 images in 1.6 seconds, while single-slice scanners are producing 1 image every 1.6 seconds. Helical scanners can be as much as 6 times faster than today’s single-slice scanners, which can be very helpful in the emergency department, in a trauma case. One could get a head-to-toe scan of a patient, or several thousand images, in less than 15 minutes.

These huge studies take considerably more time to load, display and interpret than normal examinations. Since most radiologists are paid by the number of studies on which they create reports, they will hail any technology or workflow improvement that allows them to produce a report more quickly without trading on the accuracy. Naturally, any technology or workflow alteration that requires extra time for the radiologist is bitterly resented. The solution may reside in subdividing the whole body scans using the Presentation of Grouped Procedures, which allows for the viewing of image subsets resulting from a single acquisition, where each image subset is related to a different requested procedure. Such is the case of a single, full-patient CT scan that includes images of the chest, the abdomen, and the pelvis. Even if a single acquired image set is produced, the combined use of the scheduled workflow transactions and the consistent presentation of images allows separate viewing and interpretation of the subset of images related to each of the requested procedures. Consequently, several reports from potentially different radiologists can be produced in a fashion that matches the billing policies of the healthcare facility.

The advantage with these big studies comes from an increased overall efficiency of patient care in terms of saved time and money, and from reducing the patient’s
discomfort for several trips to the hospital for additional investigation (Wendt, Peppler et al. 2002). There is also an improved quality and increased confidence of radiological diagnosis, which translates into a reduced number of calls for new tests. For example, a radiologist reading a head exam can easily access additional relevant images corresponding to the patient's neck or spine.
3 Radiology workstation Design

The radiology workstation represents the system used by radiologists for softcopy reading, and should provide support for presentation, interactive manipulation, quantification and image processing of the radiological information (Siegel and Reiner 2001). Some of the most common tasks performed during radiological interpretation include comparing images with those of previous scans, adjusting brightness and contrast, measurement of lesions, on-screen image set-up for review, image magnification, and selection of relevant images (Beard 1990). Before proceeding to a detailed description of the main workstation tools and features in Section 3.4, let us first review, for comparison purposes, the workflow for hardcopy reading.

3.1 Hardcopy interpretation workflow

For many years, the radiological images required for diagnosis, treatment planning, and treatment management were viewed on transparency film illuminated by a lightbox. In a film-based department there are two distinct alternatives to film setup, depending on the person in charge of hanging the film for interpretation.

In the first alternative, the radiologist is responsible for taking the films associated with the new study and hanging them on a series of lightboxes. Then the radiologist must look inside the film jacket for the comparison studies, and for their associated reports. After the new study is interpreted, the radiologist has to take down the films and place them into the film jacket.
In the second alternative, the film librarian is responsible for setting up the films for interpretation. Before the radiologist’s interpretation, the film librarian places the new study with any priors and reports on a mechanical film viewer, such as the multiviewer or the film alternator. The hanging of the film is done according to the specific preferences of the radiologist who will be reading the studies. By having the film librarian arrange the films the radiologist workflow is improved, but at the expense of additional time from the file room personnel. However, sometimes there are situations when the radiologist’s interpretation is delayed because studies already read have yet to be removed from the film alternator, and the new studies have yet to be hung.

A film may contain either one or two full-resolution X-ray image, or a set of cross-sectional images in different layouts, such as 4 x 5 or 5 x 6. Some multiviewers, such as the transparent band types, can move films so fast they go by in a blur. Even the slower film alternators can move a set of 4 films in about 3 s. Such speed in changing the images presented for interpretation is difficult to match by radiology workstations. In softcopy reading, the equivalent would be to have a workstation close the studies displayed for the current patient, and load all monitors with images for a different patient in only 3 s. Since a single CR image has about 10MB/image, and a single cross-sectional image has about 0.5MB/image, a workstation with 4 high-resolution monitors would have to be able to load and display about 40MB of data in just 3 s.

One of the most complex steps involved in reading conventional film is the setup of images on lightboxes or film alternators. The presentation of images for interpretation is especially challenging when prior examinations are available for review. Let us refer to the example of the MRI for thoracic degenerative disc disease, which normally has 4 series of 40 images each. Even if 20 images are printed per film, 8 sheets of film are required to print the entire study. If a prior study is also available, the radiologist would have to review 16 sheets of film. If a multiviewer with 4 lightboxes is used, the content of each lightbox must be changed 4 times in order to have all images displayed at least once. Since a series spreads across two sheets of film, only two series can be displayed at once for comparison. However, radiologists would often like to look at the two axial series from the main study, while the two axial series from the prior study are also
displayed as a reference. Because of the limited display area, the radiologist must divide such comparisons in several sequential steps, which puts a strain on his working memory. The interpretation becomes time consuming, and involves a serious amount of physical activity related to the loading and unloading of the lightboxes.

Understanding the scan patterns used by radiologists to read medical images is critical to the design of image viewing devices. For example, one study tracked the eye-gaze of four radiologists and one senior radiology resident during film interpretation of single and multiple CT examinations presented on a four-over-four lightbox (Beard 1990). Single CT exams contained 3 sheets of film with W/L optimized for displaying the mediastinum, and another 2 sheets of film with W/L optimized for displaying the lungs. During multiple CT exams reading, 1, 2 or 3 prior studies were also available.

The authors reported the following sequence of steps identified during the interpretation of a single CT exam. First, the images were removed from patient folders, sorted, and some films were placed on lightboxes. Then a systematic search, either by organ or sequentially, was performed on all images of the current study. Third, a comparison with previous studies was performed, and 3-6 critical images were reviewed in detail. Last, the report dictation took place while radiologists browsed through the study one more time.

During this experiment it was noted that film manipulation is time consuming, and that significant time was wasted on tracking the location of a particular film. Little comparison was done on film with different intensity windowed images. Radiologists constantly changed their position, moving closer or away from the film in order to zoom-in and zoom-out of the region of interest, respectively.

The sequence of steps identified during the interpretation of multiple CT exams was slightly different due to the sheer volume of film that had to be handled and read. The radiologists started by reading requisition forms and sometimes prior reports. The radiologists sorted the images after a quick scan, and piled the film sheets by intensity window per examination. Because of the large number of film sheets they had to read, the radiologists could not use their spatial knowledge of the lightbox organization to quickly
index the required image. The initial overview phase was still evident, but more diverse: the radiologists viewed the current exam and selected priors in different orders. The detail phase followed, with focus on small clusters, which were again reviewed during the final dictation. Related textual information, such as requisitions forms and radiology reports, was examined 3-5 times, which proves their great importance for selecting the interpretation strategy. Two methods were visible. In the first method, a rapid alternation occurred repeatedly between a single image in one study and an image in another study. In the second method, image clusters from two studies were compared, with fewer movements back-and-forth between clusters. Each cluster had 2 to 4 images. The smaller size of the image clusters was probably due to the film image layout, the physical restraints of keeping images close together, or attention-memory limitations.

In the next section we present for comparison the processes associated with radiological softcopy reading, followed by a few typical scenarios which simulate clinical radiology practice.

### 3.2 Viewing mode

A sequence represents a collection of related images. Usually, a sequence is synonymous with a DICOM series, but there are situations when a DICOM series can generate several sequences or pseudo-series. For example, in MRI, multiple-slices at the same location are often acquired with two different tissue contrasts. These slices are sent by the modality in the same DICOM series, which prevents efficient comparison by the radiologist. The solution is to divide the DICOM series into two sequences, corresponding to each contrast level.

For softcopy reading, the screen is divided into viewports. A viewport is a location on the screen used to display a sequence. There are two different paradigms for implementing viewports: as separate windows, which can be resized independently of the other viewports on the screen, or as matrices of disjoint, non-overlapping windows.

A sequence is usually presented on the screen in either tile or stack mode (Lou, Huang et al. 1996). In tile mode images are displayed in a grid format, similar to film. Tile mode or survey mode is generally used when reading CR and US studies, when there
is no spatial relation between images. For CT and MR examinations, tile mode causes a fragmentation of the longitudinal component of the three-dimensional information. Consequently, tile mode is used for cross-sectional imaging only to get the 'gestalt' of one particular series or of the entire examination (Beard, Molina et al. 1995). In stack mode, images are conceptually placed one on top of each other, like cards in a deck. Only the image at the top of the stack is visible. This display mode allows clinicians to create a mental 3D model of the anatomical structure in which they are interested.

### 3.2.1 Tile vs. stack

In a study comparing tile and stack modes, radiologists were asked to unravel 20 entangled tubes from the CT scan of a complex 3D phantom (Mathie and Strickland 1997). All five CT radiologists performed 3.2-5.7 times faster in stack mode (100% accuracy) than in tile mode (90% accuracy). However, a different study reported that the interpretation of the CT of bowel obstruction in stack mode was "extremely challenging and the interpretation quite time-consuming" (Memel and Berland 1995). The radiologists from this study felt that stack mode made it difficult to follow bowel loops in and out of the axial images, and to identify the transition zones and the presence of adhesions. The authors suggested these findings would apply to all structures, such as biliary ductal or urinary tract obstruction, which track through planes perpendicular or oblique to the plane of the interpreted images.

Different cross-sectional sequences can be linked together if the volumes imaged by these sequences overlap, and if all the images were acquired with the same 3D orientation. Considering a reference image in the first sequence, the application can calculate the corresponding image, acquired in the same spatial plane, in the second sequence based on the spatial information embedded in the DICOM header of these images. For example, in an MR study of the brain, an ‘Axial T2 FE’ sequence can be linked with an ‘Axial T2 SE’ sequence. These sequences can be placed in stack mode, and every time the user changes the current top of the stack image in the ‘Axial T2 FE’ sequence, the application will automatically display the corresponding image in the
‘Axial T2 SE’ sequence, taking into account the possible difference in slice thickness for two sequences.

Additional benefits of the stack mode result from the workstation’s ability to display two linked instances of the same series with different W/L settings, such as an axial series from an abdominal CT displayed with both lung and liver W/L values.

In most situations linking two sequences is straightforward and only requires the user to find one slice in each series corresponding to the same anatomical region. The situation becomes more complicated when one tries to link series with non-parallel axis, and some applications prohibit such attempts in situations like the one illustrated in Figure 26.

![Figure 26: Sequences with non-parallel axis](image)

Stack mode is more commonly used than tile mode for CT/MR, because of the lack of screen space compared to the viewbox. Also in tile mode it is difficult to properly register images from current and previous studies. Another negative aspect related to tile mode comes from the ‘information overload’ that makes it difficult for radiologists to pay attention to all tiled images. A 1994 study involving 3 experienced radiologists and 10 synthesized MR studies showed an improvement in efficiency with linked stack reading 1.5-4 times faster than film (Chang and Ziegelbein 1994b). Another study from 1994 involving 5 radiologists and 20 synthesized MR studies with priors resulted in both efficiency (twice as fast as film) and accuracy improvement (Chang and Ziegelbein 1994a). Similar results for reading MR examinations showed an average time of 58 seconds using a linked stack mode vs. 80 seconds for the tile mode (Niimi, Shimamoto et al. 1997). A 1995 study showed that the stack mode improved the radiologists’ ability to detect nodules in spiral CT of the chest, and a follow-up study from 2000 indicated that
the stack mode should be used to "clarify the adjacent vascular anatomy of perceived nodules" (Seltzer, Judy et al. 1995; Jacobson, Judy et al. 2000).

It has been showed that shapes can be depicted and recognized using different attributes, such as grayscale, color, texture, relative motion and depth from binocular disparity, which correspond to discrete information-processing channels in the human visual system (Bertin 1981). Stack mode may activate additional perception attributes and improve lesion detection (Seltzer, Judy et al. 1994).

To summarize, stack mode is more appropriate to use than tile mode for:
- tracing longitudinal structures, such as blood vessels, muscles, nerves, and tubes
- assessing surface contour of organs, such as liver and kidney
- separation of adjacent structures, such as lymph nodes from blood vessels.

However, tile mode is still seen as an alternative viewing paradigm. One reason for the use of the tile mode comes from the radiologist's training habits. Since most radiologists are trained using film, the way to avoid developing a new viewing strategy is to use the 'digital light-boxes' metaphor, with images displayed in tile mode. The other important reason for choosing tile instead of stack is the lack of detail and context in stack mode. Since only one image is displayed at any time in stack mode, it is sometimes difficult to mentally visualize a 3D section of the anatomy. One study explored an alternate way to display both detail and context in a tiled display through a novel zooming technique, and concluded that users' performance was comparable to that using thumbnails (Kuederle, Atkins et al. 2001). Preliminary results from a study comparing stack mode with this enhance tile mode that provides detail-in-context indicate that for certain tasks, stack mode can be twice as fast as the tiled approach (Hunter, Susanto et al. 2001). It is possible that an augmented stack mode with some context around the focal image might provide the best of both worlds.
3.3 Radiological softcopy workflow: task decomposition

In a landmark study of hardcopy reading, Rogers et al. found the reading process can be divided into three steps (Rogers, Johnston et al. 1985):

- general information gathering, including review of the patient’s history, radiological examination history, along with a quick sort through the films of the current study
- rapid view of the image set to detect candidate pathological states
- the actual ‘reading’ or diagnostic phase, when the radiologist’s attention is focused on those candidate abnormalities, with the dismissal of some as non-pathological and recognition of others as significant

A similar structure of the softcopy reading task is presented by Stewart, and includes the following steps (Stewart 1993):

- Case selection, from a study list or by searching for a patient name.
- Survey or initial screening phase, characterised by a rapid view of all images and corresponding priors. Tile is the default display mode for new cross-sectional exams. During this interpretation step, the radiologist may select a few images that show patient pathology.
- Case presentation step, where auxiliary text information, such as demographics, case history, and previous reports, is used to orient the user of the suspected patient problem.
- Case interpretation step, where several image processing tools may be used, as detailed in the next section.
- Documentation of diagnosis: unambiguous, accurate, concise, and accountable report; use graphic pointers and voice report playback.
- Presentation of diagnosis: concise structured report.

At the SPIE Medical Imaging conference from February 2003, several radiologists provided an interactive insight in the processes of softcopy interpretation during a workshop called ‘Radiologist interpretation of medical images - How they do it’. The three participants were Dr. Ronald Summers from the National Institutes of Health, and Dr. Steve Horii and Dr. Harold Kundel from University of Pennsylvania.

The first study reviewed by these 3 radiologists was a case of moderately-subtle CR of lung cancer, with no interactive image manipulation allowed. Dr. Kundel reported he reads such cases in 3 steps, corresponding to the first impression, a systematic check,
and dictation. Dr. Summers reported the following interpretation steps for softcopy reading of chest CR:

- preliminary step dedicated to quality assurance, when he checks the films match the requisition order, confirms the film orientation and the image quality (film exposure)
- gestalt step, when he quickly examines the image while ignoring abnormalities
- systematic search, starting at top of the image, then moving back and forth between ribs, to finish in the mediastinum area.

For CR reading, Dr. Horii reported the following, similar interpretation steps:

- preliminary quality assurance step
- gestalt, or the overall impression step
- visual search covering all corners of the image, intended to detect the major abnormalities
- fourth step for searching the areas known for missed abnormalities.

The second study reviewed by these 3 radiologists was a case of CR emphysema, with only the posterior-anterior projection shown. Dr. Kundel, who reported using the W/L, the magic glass\(^5\) and greyscale image invert\(^6\), paces himself to read about 20 chest exams per hour. Dr. Summers also reported using the magnification quite often, pacing himself to read 15-20 chest exams per hour. Dr. Horii, who reported using mostly the invert tool because he finds W/L adjustments to be time consuming, also reads about 20 chest exams per hour.

The third study reviewed during the workshop was a case of CT chest. Dr. Summers noted that using tile mode for softcopy reading involves too much eye movement, and this repetitive stress of looking at each image 3 times, with different W/L settings, would be very annoying. Dr. Summers uses a single predefined W/L for lung, combined with a methodical scan in stack mode and an adjustable scroll speed. He finds fast scroll not to be very useful. Dr. Horii starts with a quick scan for abnormalities in stack mode. If he sees any abnormalities, he sets a mental reminder using anatomical

---

\(^{5}\) A user defined, movable, rectangular or elliptical area that can be manipulated independently from the rest of the image.

\(^{6}\) Inverts the display function map so as to display bone as black.

*Designing Better User Interfaces for Radiology Workstations*
landmarks, and not by using the image number or the 'Flag key images' feature. He also prefers to use the stack mode, as it 'lets you follow the vessel, and things pop-out'.

We distinguished the following phases of a typical softcopy interpretation process. These phases are described in detail in the next section:

1. radiologist logs-in
2. a study, called the main exam, is selected for interpretation
3. the images from the main exam are loaded by the workstation and displayed on screen
4. whenever available, prior examinations may be also loaded on the workstation for review, together with their associated radiological reports
5. radiologist interprets the main exam
6. radiologist dictates the report on the main exam
7. steps 2-6 are repeated until there are no more studies for interpretation, or until the radiologist decides to stop reviewing studies
8. radiologist logs-out from the system

The radiologist log-in serves to authenticate the user into the system. A personalized, user-specific profile is attached to each user, depending on their role and their preferences, such as worklists and hanging protocols (see below).

After logging in, the diagnostic software is launched automatically and (preferably) a user-specific or group-specific 'worklist' is displayed for the main study selection. Worklists are used to define the type of unread studies that are to be presented for interpretation and their order of presentation. The worklist is a filter into the PACS database of studies yet to be interpreted. A worklist uses the model of the film jacket, having a hierarchical organization with patient, study, series, and image entities. Thus a radiologist can read studies from a worklist limited to a certain modalities, such as CT and MRI, and to a certain anatomic area, such as brain and spine.

The images associated with the main study are usually stored on the hard-discs located on the image servers. These servers are connected to the workstations using fast networks with 100Mb/s or 1GB/s transfer rates. In the classical environment where
images are available on demand, the workstations should provide a quick response time between when the radiologist triggered the opening of the main study, until the first images are displayed on screen. An acceptable response time is less than 2 s, and involves either the display of a large image, such a CR image with 10MB of data, or the display of 6 smaller cross-sectional images from CT or MRI.

The arrangement of images immediately after opening a study is governed by a Hanging Protocol (HP). A HP specifies how the available screen real estate is divided into viewports, what series is displayed in each viewport, and the presentation attributes of each series, including window width and level, zoom and pan, tile or stack mode. Whenever available, relevant prior exams and their reports should be loaded for review. Usually the relevant priors belong to the same modality as the main exam and they are used for historical comparison. It is also possible for the relevant priors to belong to a different modality than the main exam and they are used for inter-study correlation.

The exam interpretation phase can involve several steps, depending on the modality and radiological procedure the main exam belongs to, as presented above. The same series of images can be viewed multiple times, under different presentations, such is often the case of a CT of the abdomen that can be reviewed with lung and liver W/L settings.

The dictation phase concludes with the presentation of the diagnosis, and the creation of a concise structured report. In a traditional voice dictation environment, the spelling and formatting of the report is done by the transcriptionist. At a latter time the radiologist has to review and sign the report. It is estimated that healthcare institutions spend on radiology transcription $600 million to $1.2 billion annually (Page 2000).

If voice recognition (VR) software is used, the radiologist performs the spelling, formatting, and digital signing of the report immediately following the interpretation phase. Key benefits of VR are reduced transcription cost and report turnaround time (Seltzer, Kelly et al. 1997). For example, at Massachusetts General Hospital, report turnaround time has gone from 5 days to less than two with VR, while saving $750,000 by eliminating the radiology transcription cost. Since radiologists view VR reports
directly on their PCs, the quality of reports may increase because visual cues can act as reminders to include additional information or modify specific sections of the dictation (Antiles, Couris et al. 2000). However, some radiologists believe VR systems (with accuracy rate of about 95-98%) take more of their time than traditional transcription methods did. Consequently, these radiologists wonder why hospital administrators want their highest paid and highest revenue-generating staff spending time unproductively (Page 2000). This thesis is not concerned with the dictation phase, as this is a whole research topic.

Once the dictation is completed, the radiologist closes the exam as 'dictated,' which should automatically remove the exam from the worklist of un-reported exams. A radiology workstation could have an 'automatically load next exam' feature, which will take care of loading the next exam, thus making phase 2 transparent for the radiologist. The selection of the next exam from the worklist should be done automatically according to sorting criteria defined by radiologists. Examples of such criteria are: alphabetical on the patient’s name, and chronological on the exam time. In a hospital environment, a priority scheme should supersede the default sorting criteria, allowing for ‘STAT’ (urgent) exams to be reviewed first.

It is possible for worklists to become empty. Since no study can be selected for automatic loading, the workstation screen(s) should be empty, and an informative message should be displayed, such as "The current worklist is empty." In such event, the radiologist may decide to relax the filtering used for exam inclusion in the worklist. Alternatively, the radiologist could wait for the next eligible study to appear in the worklist. In order to avoid the radiologist periodically checking if a new exam has appeared in the worklist, the workstation should automatically load the first new study arrived in the worklist, and eventually display a message like “A new exam arrived at <time> and was automatically loaded.”

3.4 Radiology workstation tools and features

A medical imaging workstation is a computer graphics-intensive device used by health care professionals to interactively display and manipulate images (Kim, Fahy et al.
1988; Arenson, Chakraborty et al. 1990; Ho, Ratib et al. 1991). A radiology workstation is primarily designed to display, analyze, and process greyscale images. The radiology workstation must support a high greyscale resolution of 8 bits or more, and must include a significant variety of image analysis and processing capabilities.

The radiology workstation must support these operations in an interactive way, without excessive delays (Kim, Fahy et al. 1988). Beard et al. also stressed the importance of a low response time for frequently used tools, such as W/L, zoom/pan, and scroll (Beard 1991). Consistency was considered possibly more critical than absolute response time, once that time dropped under a certain threshold. An intelligent workstation should be adaptive to the user’s preferences, fast, and able to handle both simple and complex tasks (Stewart, Aberle et al. 1993).

In order to extract all the necessary clinically useful information from the radiological examinations, several steps are required, and each step may require the use of different workstation tools. Some of the required workstation functionality was presented as early as 1990, and included (Beard 1990):

- an image index for rapid access to images, folders, and exams;
- sequential and arbitrary image display; access to all text information;
- ability to display at least 9 full-resolution images;
- ability to compare different image clusters;
- W/L and zoom.

Stewart et al. grouped the workstation tools based on their role, and defined the following tasks: case selection, case presentation, and case interpretation, which were presented in Section 3.2 (Stewart, Aberle et al. 1993). During the case interpretation step, the following tools may be used:

- Interactive manipulation: w/l, zoom/pan, measure/annotate
- Image processing: image subtraction, data compression, edge enhancement, volumetric displays, contour extraction. These tools are rarely used either because they are time consuming, or because the radiologist is not aware of their utility.
- Quantification functions: measurements, histogram analysis, texture analysis.
Horii presented a classification and a description\(^7\) of the workstation functions and features that included navigation functions, image manipulation functions, and image management functions (Horii 2002). According to Horii, navigation refers to the task of “finding the examination to be reviewed, and a specific examination when there is more than one for a patient.” By radiology workstation navigation we understand the workstation metaphor, and the commands and hand motions used to access patient folders and to display images. ‘The worklist’, ‘List all’, ‘Folder display’, ‘Next patient’, ‘Previous patient’, ‘Next exam’, ‘Previous exam’, ‘Resume last’, and ‘Mark as read’ are some examples of navigation functions. Examples of image manipulation functions include W/L, greyscale invert, zoom/pan, flip/rotate, measure and annotate. Image management functions include ‘Delete exam’, ‘Mark for teaching’, ‘Mark for non-deletion’, ‘Print’, ‘Hanging protocols’, and ‘In progress examinations’.

We propose a more refined classification of the workstation tools into four categories. Our first category is called Study management tools, and includes the navigation tools described above. We suggest subdividing the Study management category to include tools for:

- study selection: worklist, all unreported exams, and study finder
- study closing: close study as either reported, or unreported, or reviewed
- study context switching: next / previous study, next / previous patient, automatic opening next exam

Our second category includes tools used to optimize the presentation of images for study reading and reporting. These tools are often included in the image processing pipeline implemented by the workstation software, and include W/L, greyscale invert, zoom/pan, flip/rotate, shutter, display of patient demographic info and annotations. Other image enhancement tools include edge enhancement, smoothing, greyscale histogram equalization, and image arithmetic (adding or subtracting images). This category would also include the measurement and annotation tools.

---

\(^7\) We are not presenting a description for these functions, as this is done in the reference provided.
Our third category includes the tools used to organize the information on the computer screens. These tools are used to arrange the images for easy registration and comparison. We can further subdivide this category to include tools:

- affecting the screen division into one or more viewports, such as changing the layout (for example from 2 rows by 3 columns, to 4 rows by 5 columns) and selecting the monitors used to display a series;
- affecting a series associated with a particular viewport, such as changing the image display mode (for example, from tile mode to stack mode) and linking two series together for synchronized navigation.

We can also classify the workstation tools based on their scope. Consequently, there are tools operating at the patient level, such as closing a patient, or switching between two patients. Other tools operate at the study level, such as marking a study as reported, printing a study, or sending a study to another DICOM application entity. Other tools operate at the series level, such as selecting the tile or stack display mode, or applying a predefined W/L preset to an entire stack of images. Finally, there are tools that are image specific, such as placing an annotation marker, or measuring the linear size of a lesion. Note however that some tools can operate under different scopes, so the workstation should provide operators to set the desired scope. For example, one could place an annotation on a single image, or adjust the window width and level for a single image, which is the case for images from modalities like CR or US. For cross-sectional imaging, images are usually presented in stack mode, and radiologists usually like to change the W/L of the entire series to a certain preset, such as ‘bone’ or ‘soft tissue’. Also, for CT or MRI of the lumbar spine, it is important to have a spine labelling tool. Such a tool works by first having the radiologist label one image from a sagittal series, and then the workstation would apply the same labels in every image from that stack, as illustrated in Section 2.1.5.1.

Several papers reported that W/L adjustment is the most frequently used tool, and our user study from November 1999 confirmed this usage pattern (Honeyman, Arenson et al. 1993; Horii and Kundel 1994; Moise and Atkins 2002c). Other commonly used tools are greyscale invert, magnification and flip/rotate tools. However, some physicians
avoid using these tools, as they increase the complexity of interaction (Beard, Johnston et al. 1990).

3.5 Workstation tools usage patterns

One study investigated the workstation tool usage patterns of radiologists and ICU physicians for reading portable chest radiographs (Horii, Feingold et al. 1996). No image manipulation operation was performed for about two-thirds of the studies. Adjustment of W/L had the highest percentage of use overall, followed by zoom, video invert, and high resolution. In this study radiologists used W/L adjustments to a greater degree than ICU physicians. However, the use of other image manipulation functions by radiologists was considerably lower: zoom, invert greyscale, and high-resolution display were not used at all.

It was found that the use of workstation tools by radiologists changes with increasing experience with the system. For example, radiologists tend to use tools such as image zoom and magnify less frequently as they gain experience with the system (Siegel 2002). However, even experienced radiologists utilize W/L adjustments in the majority of cases.

Such patterns of tools utilizations may seem a paradox at first: one may expect the more familiar a user becomes with a system, the more likely he or she will be to take advantage of additional functionality. One may assume that if the latter situation does not occur, it is because the extra tools are cumbersome to use. By interviewing several experienced radiologists and expert users, we learned they avoid using even simple to use tools, such as flagging key images. However, image flagging can be implemented as simply as pressing the ‘Space’ bar, which is quick and requires no user eye focus. Even if tools such as image flagging can have such an ergonomic implementation, there is still a question of how easy it is for the radiologist to navigate to and visualize these key images.

We believe there are several other factors that contribute to these workstation utilization patterns, which we briefly explain below. First, there is an effect of the user gradually getting comfortable with the softcopy interpretation task. The physician
progressively increases his or her confidence that the workstation produces an appropriate rendering of the medical images for interpretation. Second, there is an issue of keeping a reasonably fast pace for interpretation. Each radiologist finds he or she performs better at a certain tempo, optimally adapted to his or her visual acuity and cognitive resources. As long as the radiologist maintains this pace, his or her productivity is high, by being able to produce reports faster and with higher accuracy. The tendency to work towards a short interpretation time has financial connotations. Also, shorter interpretation times lead to a lower probability of radiologists being interrupted while half-way through the interpretation of a complex case, in order to perform a secondary task such as providing consultation for a referring physician. Such interruptions represent significant disruptions of the reading workflow, because all the information stored in the radiologist's working memory is flushed out, and the entire clinical context of the interpretation is lost. After resuming from the interruption, the radiologist has to bring back the relevant information into the working memory. Consequently, the radiologist has to review some of the patient's relevant information (mostly images, but also textual information such as previous radiology reports) either to identify the key findings one more time, or to retrieve the tips (pointers) required to find this information in his or her long term memory. And third, there is a trade-off between the use of additional tools and the disruption of the reading process. Several studies identified the expert reader's ability to compensate for difficult reading conditions, such as poorly acquired images, low-resolution or de-calibrated monitors, and small size images (Seltzer, Judy et al. 1998; Parr, Anderson et al. 2001). Consequently, there are often situations when using additional tools does not change the outcome of the interpretation, and only the radiologist's confidence may increase. However, when the radiologist's confidence in the interpretation is high enough, the use of additional workstation tools is not merited due to the associated time penalty and disruptions of the reading task.

3.6 Classification of PACS workstations based on their role

Expressions like PACS workstation, digital workstation, diagnostic workstation and radiology workstation are sometimes used interchangeably to define an environment
for softcopy presentation. The type of imaging workstation dependents strongly on the role and objective of its primary user.

According to one classification, there is a mono/multi modality reporting station, a viewing workstation, and an image processing workstation (Dwyer 2000). A mono-modality workstation is usually associated with a mini-PACS. A viewing workstation is used for a fast reference system by the clinician, while an image processing workstation generally includes sophisticated high-end systems, with specialized hardware and dedicated software packages for use by the medical staff.

A second classification is based on primary role, functionality and hardware configuration, and includes primary diagnostic workstations, secondary review stations and tertiary workstations, as described below (Matt and Dennis 1999).

### 3.6.1 Primary diagnostic workstation

The primary diagnostic workstation is used by radiologists to generate diagnostic reports, so it should be designed for productivity and diagnostic accuracy. The radiologist’s job is to observe and recognize abnormal visual patterns that may not even have been looked for. The basic tools for this task include zoom/pan, W/L adjustments, comparison with previous studies, measurement and annotations. More advanced tools will allow for customizable hanging protocols, study bookmark or parking, linking series, automatic loading of the next study, 3D reconstruction and interactive collaboration. The configuration of the radiology workstation should be flexible to allow user profiles based on the user’s role and the type of data reviewed.

Workstations with two greyscale, high-resolution (2500x2000 pixels, or 2000x1500 pixels) monitors are the most common configuration, but workstations with four such monitors are also available. There is also a new trend to extend the screen display area with an additional colour monitor that could be used for:

- Displaying colour images from nuclear medicine or Doppler ultrasound
- Displaying text-based information, such as previous radiology reports, as reading text from the high-resolution monitors is inconvenient for the radiologists
Separating diagnostic presentation and navigation contexts, by providing a navigation console for setting up the entire screen presentation. Distracting menus and icons can be removed from the quality diagnostic monitors, which will solely be used for image display and interpretation thereby increasing image quality and overall reading efficiency. A scaled-down version of the content for each greyscale monitor is displayed on the colour monitor, which becomes a virtual desktop. Image presentation on all the greyscale monitors can be set up solely by operating the workstation controls on the colour monitor, which has a smaller physical size and thus requires less mouse travel. The colour monitor also provides the workstation user with a gestalt of the information presented on all the monitors.

3.6.2 Secondary review workstation

In general, referring physicians rely on reports provided by radiologists, but other medical personnel make patient care decisions by directly viewing radiological images. For example, high-end secondary review workstations displaying chest radiographs are used by physicians in the ICU to adjust the position of tubes and lines placed inside a patient. An orthopaedist can act on a fracture based on radiographs displayed on a secondary review workstation with two greyscale monitors with a spatial resolution of 1600x1200 pixels.

These workstations will provide faster access to STAT (urgent) studies and reports than is possible in a film-based environment, enabling real-time interaction between clinicians and radiologists, and curtailing unproductive travel for clinicians to the radiology department for collaboration (Welz, Ligier et al. 1995; Bellon, Wauter et al. 1997).

3.6.3 Tertiary workstation

The tertiary workstation represents a low-cost alternative for the distribution of reports and images to referring physicians, on-call radiologists or for teaching purposes. In order to operate under a lower connection bandwidth, compression of images is sometimes required. Virtual private network is the most common solution for protecting the confidentiality of the images and other patient information.
Any modern desktop or desktop-replacement notebook can function as a tertiary workstation, and colour displays with a resolution of 1280x1024 pixels are usually acceptable.

### 3.7 Design

Most design principles enumerated in this section are based on the author’s informed opinion and hands-on experience with the design of state-of-the-art radiology workstations. Some of these principles have not yet been validated by research.

Since the radiology workstations are the sole interfaces of radiologists to PACS, a proper user-centred design is important. The evolution of modern radiology workstations is marked by significant improvements from the previous generation, as described in the next section. These advances are due to more complete requirements, better understanding of the medical imaging field, and improved software development policies that incorporates more feedback from the end users (Erickson and Kossak 2000).

The workstation software should combine general design principles with domain-specific requirements, based on a deep understanding of the radiological visual scanning and diagnostic workflow. The workstation should volunteer and organize the information, and should provide context-sensitive tool selection to support natural pathways of human reasoning and classification. The design of the workstation software should be based on a deep understanding of the radiological visual scanning patterns and the sequence of operations commonly performed during the diagnostic process. The workstation should provide sensory feedback, placing items in search area where they are most likely expected. The user interface should be consistent and should require minimal memorization. The workstation should provide, organize and present the information so that it may be comprehended easily.

As general principles we suggest a user interface requiring minimal memorization, providing sensory feedback, and placing items in search areas where they are most likely expected. For a fast, ‘streamlined’ interpretation of the images, one should be able to customize the layout of the controls (Horii, Grevera et al. 1998). The main toolbar should only contain the icons for the most frequently used software tools, to
reduce screen clutter and to simplify tool selection. The tools that are used infrequently should be made available only on a per-need basis, even if this would involve more interaction from the user to activate them. The user should be able to specify the screen location of the main toolbar, to accommodate different screen formats: top/bottom of the screen for portrait monitors, left/right of the screen for landscape monitors.

Most modern radiology workstations now provide similar functionality. We now witness a shift in the priorities of PACS vendors, from adding more features onto emphasis on interactive aspects. Prospective customers will stop asking the question “who has the most features” and replace it with “which workstation best fits our workflow.” A high performance implementation of each workstation feature, without concern for the integration, will likely produce local optimal functionality, with sub optimal efficiency in the overall completion of the radiology task.

The radiology workstation should target the natural completion of tasks specific to radiological practice and should adapt to user specific needs. The digital system should adapt to the specific working conditions and should provide the contextual information available in the older film-based environment. For example, in the intensive care unit a patient-bed relationship should be possible, and also the placement of the prior image adjacent to the current image (Reiner, Siegel et al. 1999).

Instead of the user having several computer stations, each one with its own operating system, user interface and limitations, only a single access point should remain. This would need a tighter integration between the vendors (‘plug-and-play’ PACS technology) in the context of the extension of the current standards and guidelines. Web-based applications and easy-to-download thin clients could be used, with on-demand loading of the appropriate modules based on the user’s role and type of data accessed. On-the-fly activation of the latest version of the software from the central server would allow for easy software update, maintenance and support.

3.8 Evolutionary stages of radiology workstation

The first radiology workstations were produced in the early 1990s, either by adding more features to a mini-workstation, or by presenting prototypes for validation to
representative users. These first attempts mostly represented an engineer’s rather clumsy solution to a problem poorly defined due to lack of experience and understanding of radiologists’ workflow. Mostly an awkward ‘one monitor, one-image-per-monitor’ display paradigm was used (Erickson 1997).

The second evolutionary step took place in the late 1990s, and it was characterized by porting the initial prototypes from platforms like UNIX, Next or Macintosh into the ‘Wintel’ mainstream (Ernst, Le et al. 1999; Kolodny, Raptopolous et al. 1999; Siegel and Reiner 2001). Most workstations in this stage used a slow DICOM query mechanism to get the images into the local cache. Some of the new features included sequential image playback, W/L presets, magnifying glass and support for basic measurements.

Modern radiology workstations represent a significant improvement in speed and functionality from the previous generation. These advances are due to more complete requirements, a better understanding of the medical imaging field and improved software development policy that incorporates more feedback from the end users. In order to expand the potential of a digital hospital, re-engineering the radiology processes resulted in significant changes in the workflow. However, despite all this progress, this stage is still marked by the lack of naturally integrated navigation and manipulation tools.

3.9 Identifying the representative users

One good development policy is to present early prototypes to as many representative users as possible (Coble, Maffitt et al. 1995). This development policy relies on research in usability engineering fields, such as usability testing and rapid prototyping. Usability testing refers to the evaluation of information systems that involves participants who are representative of the target user population (Kushniruk, Patel et al. 1997). Rapid prototyping represents a design methodology based on building successive demonstrator systems (Bellon, Feron et al. 1996a). These demonstrators are designed to express the initial ideas in a concrete form, and are kept sufficiently simple to allow rapid changes. What an engineer may think is an optimal solution because it is ‘technically better’ can be considered a sub-optimal solution by the radiologist.
Representative users can be classified based on several categories, which are described below. A more detailed analysis and discussion on the patterns of tools utilization associated with these categories is presented in Section 3.6.

### 3.9.1 Role and hospital department

A user’s role, usually reflected by the hospital department he or she belongs to, is what determines the main purpose of a PACS workstation. Accordingly, the workstation should provide the flexibility to accommodate users from various hospital departments such as radiology, emergency and surgery.

In a study comparing the use of workstations for reading portable chest radiographs, radiologists used window and level adjustment more often than the ICU physicians (Horii, Feingold et al. 1996). However, in the same study it was reported that ICU physicians used image manipulation functions such as zoom, invert, and full-resolution display, more often than the radiologists.

Several workstation configuration profiles should be available to deal with desktop appearance, system functionality, and layout of controls associated with various user roles. Some radiology workstation features should only be available to radiologists, some only to surgeons, and some only to physicians from the ICU (Pomerantz, Siegel et al. 1996).

For example, report creation and marking studies as reported should only be available to radiologists. Referring physicians should have some limitations on certain features, such as being able to annotate images, but these annotations will not be saved by PACS when the study is closed. Deleting images and moving studies between patients should be available to PACS administrators only.

In another example, the ICU may require a high-resolution, single-monitor diagnostic workstation with a simplified set of tools and hanging protocols, while the radiology department may have up to four high-resolution monitors attached to a workstation, with a complex hanging protocols setup and the full range of tools and features.
For a big hospital it may be very important to provide secure logging to radiologists and referring physicians, to have several users involved in an interactive collaboration, or to have a pager/email based system for notification (Mack, Holstein et al. 2000) on availability of reports. For a university hospital the access to some core capabilities of the workstation through plugs-in it may prove to be of great value. A small group of clinicians or other at-home readers will appreciate the ability to ‘sell’ overflow work or on-call services.

3.9.2 Training level

There are several levels of medical competence, corresponding to different levels of training. Users of the radiology workstation can be classified as faculty, residents, fellows and students. Differences in memory, knowledge or skills can affect medical decision-making. One study showed a correlation between experience and a decrease in the discrepancy rates in the interpretation of CT scans, with the trainees more likely to miss findings than to read normal scans as abnormal (Wechsler, Spettell et al. 1996). Different tools, such as our extension to hanging protocols presented in the next chapter, can be used to help novice users to learn the reading patterns of the experts.

The expert has attained a high level of standing in the medical domain, such as a board certification. The expert’s superior perception of anatomical patterns is mediated by highly structured and richly interconnected domain-specific knowledge. It is estimated that it takes about 10 years of devoted effort to acquire the approximately 50,000 chunks of information stored in the long term memory of an expert. As a result, experts can monitor their problem solving, allocating time appropriately by noting their errors and checking questionable solutions (Pantel, Glaser et al. 2000).

For the intermediate user, periods of smooth incremental learning may be followed by periods of consolidation of knowledge that may cause a decline in performance. The intermediate user is likely to spend a substantial time searching for information because the extensive body of knowledge that was just acquired has not been organized yet. Intermediates provide unnecessary elaborations in explaining patient
problems, elicit considerable amounts of information from the patient to make a
diagnosis or request unnecessary additional tests.

Novices, on the other hand, do not conduct irrelevant searches, simply because
their knowledge is not rich enough. A novice's knowledge base is sparse and an expert's
knowledge base is intricately connected, whereas an intermediate may have a lot of the
pieces of knowledge in place but lacks the extensive connectedness of the expert (Pantel,

3.9.3 Level of experience with radiology workstations

Based on their experience with computers, PACS and more importantly radiology
workstations, we will find novice, intermediate and expert users. It was found that the use
of workstation tools by radiologists changes with increasing experience with the system.
For example, the radiologists tend to use tools, such as image zoom and magnify, less
frequently as they gain experience with the system (Siegel and Reiner 2002b).

When rating a radiology workstation, novice users will look at:

- how intuitive the controls are,
- how easy it is to learn the basic functionality required to perform the main task,
  which is the creation of diagnostic reports by radiologists,
- how easy is to find guidance in the online help or the user's manual,
- how good is the default configuration, including the tools available on the main
toolbar, the size and the type of the fonts, and the window level presets.

Expert users will look for short cuts and hot-keys to maximize their efficiency, to
give them access to the unleashed functionality of the workstation. In Table 1 we
synthesized several requirements for workstation controls for novice and expert users.

Commenting on the availability of appropriate controls, Dr. Steven Horii, a
radiologist and a power-user, was writing about how important is to:

- have the window width and level presets and any manipulation tools
  (zoom, roam, etc.) available quickly as needed. Not all the vendors can do
  this (Horii 2003b).
Table 1: A comparison between expert and novice users

<table>
<thead>
<tr>
<th></th>
<th>Novice</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals</td>
<td>Be able to learn how to perform the main task (report creation)</td>
<td>Maximize efficiency</td>
</tr>
<tr>
<td>Ergonomics</td>
<td>Intuitive, easy to learn, and easy to find guidance</td>
<td>Heads-up, streamlined interpretation</td>
</tr>
<tr>
<td>Configuration</td>
<td>Default presets</td>
<td>Personalised</td>
</tr>
<tr>
<td>UI requirements</td>
<td>Essential tools one click away, either available as icons on the main toolbar, or in context sensitive menus</td>
<td>Short-cuts and hot-keys</td>
</tr>
<tr>
<td></td>
<td>Default hanging protocols</td>
<td>Pseudo-modes</td>
</tr>
<tr>
<td></td>
<td>Automatic loading of relevant prior examinations</td>
<td>User-specific mapping of controls</td>
</tr>
</tbody>
</table>

3.9.4 Organization

The radiology workstation can be used in different organizations, from large, multi-site hospitals with a big radiology department, to medium sized radiology clinics employing non-specialized radiologists, or even a small group of radiologists serving small medical imaging centres.

3.9.5 Area of expertise (specialty)

Radiologists can be specialised in one or more of these disciplines: X-Ray Radiologist, CT/MR (cross-sectional) Radiologist, Sonologist, NM Radiologist, Neuroradiologist, Interventional Radiologist.

The data coming from different imaging modalities is varied, with image matrices varying from 64 x 64 pixels in NM to 4096 x 6144 pixels in mammography, from colour US and pseudo-colours in NM to greyscale in CT/MR/X-Ray, and from uniform groups of cross-sectional images in CT/MR to unrelated, heterogeneous images for the other modalities. The needs of the radiologists are highly dependent on the kind of examinations they view, so the workstation tools should also be modality-specific.
For example a NM Radiologist may be interested in applying various pseudo-colour palettes to a greyscale image. However, for a CT/MR radiologist such tool will usually be of no value, but he/she will appreciate the tools corresponding to cross-sectional modalities, like stacking and linking different series together. For a Sonologist it will be very important to manipulate (play, trim, annotate) cine images. An X-Ray Radiologist will do a lot of ‘Panning and Zooming’ and may also use image processing tools like filters for edge detection.

3.9.6 Body part

Specific needs of a radiologist can also be associated with the anatomy they are viewing. For example, a radiologist specialized in body extremities will be interested in a tool that allows one to superimpose and adjust different prosthetic implants on the X-Ray of someone’s knee or hip. A radiologist specialized in chest X-ray, bone X-ray, or mammography will often use a ‘Zoom’ to take a closer look at various anatomical locations with a high probability of hiding lung nodules.

A radiologist reading a CT spine will be interested in a ‘Vertebrae labelling tool’ that will allow for an easy placement of annotation labels on the vertebrae. The user will only label the vertebrae from the top-most image in a stack, as shown in Figure 27. The system will then automatically apply the corresponding labels to all images in that stack, as illustrated in Figure 28 and Figure 29.
Three-dimensional reconstruction tools, such as maximum intensity projection and multi-planar reformatting, could be handy for a radiologist interested in examining the blood vessels from an magnetic resonance angiography exam. A radiologist reading CT/MR of the main body will appreciate a three-dimensional localizer, which allows the user to select a particular location on an axial image, and then identifies the corresponding location on images acquired in the coronal and sagittal plane. An example of localization in 3D is shown in Figure 30.
3.10 Case Study: Radiology workstation design for the ICU

The ‘one-size-fits-all’ approach for radiology workstation design is not good enough anymore. While most of the PACS vendors are racing to add more features to the radiology workstation, there is little interest in addressing the specific needs of other hospital departments. This section addresses the main objectives in designing a radiology workstation for use in ICU.

3.10.1 Specific conditions in the ICU

ICU departments have a grouping of beds for patients requiring critical care. These patients require intensive monitoring. Non-routine, problem-specific chest radiographs comprise 25%-40% of all ICU examinations (Shile, Kundel et al. 1996).

ICU physicians are asked to make rapid decisions, based on limited and usually incomplete information on patients who are strangers, with no prior medical information available to help with the diagnosis. ICU physicians require more rapid access to images and radiology interpretations than in most hospital settings.

PACS can provide faster, more convenient access to radiological images than film does. For example, PACS can shorten the time elapsed until the images are reviewed by ICU physician by over 30% (Andriole, Storto et al. 1996; Horii, Kundel et al. 2000). PACS can also reduce the number of examinations per patient day from 1.3 to 1.09, and decrease the time to clinical action from 201 min to 126 min (Andriole, Storto et al. 1996).

Convenient clinician review of radiological images can save an average of 44 minutes daily (Reiner, Siegel et al. 1998a). However, the introduction of PACS can also reduce the communication between the physician and the radiologist by 82% for conventional radiography, and 44% for cross-sectional imaging (Reiner, Siegel et al. 1999).

Despite these reported improvements in the report turnaround time, the delay from image acquisition until report transcription is still too long for the ICU specific requirements. Consequently, running to the radiology department to get the interpretation
on a critical examination is still the primary means by which ICU staff can get fast exam interpretation.

**3.10.2 ICU specific features for the PACS workstation**

The patients currently in the ICU ward can be listed using their names, or their bed number. If tracking the patient’s position in the ICU is not done automatically, bed initialization functions such as *Move*, *Clear* and *Init* should be implemented (Nahmias, Kenyon et al. 1997). When idle, the radiology workstation may display a gestalt of the current patient situation, such as a thumbnail image of the last examination for every patient. From this ‘emergency review’ or ‘View all’ mode (a 4x5 layout of 256x256 pixels thumbnails), any patient can be selected by name or by bed number (Bellon, Feron et al. 1996b); (Nahmias, Kenyon et al. 1997). For the ‘historical review’ of the selected patient, ICU specific hanging protocols should be used to minimize the time spent rearranging images on monitors (Bellon, Feron et al. 1996b); (Moise and Atkins 2002c); (Strickland and Allison 1995). For example, the most recent examination should be displayed in real-size on the left monitor, and the last 4 priors should be displayed four up on the right screen in reverse chronological order; other papers suggested 6 and respectively 16 ‘iconified’ priors (Strickland 2000), (Nahmias, Kenyon et al. 1997). A double-click with the mouse could be used on any prior images to switch back and forth between real size and the four up layout. A ‘daily overview’ mode, listing for comparison the two most recent exams for every patient, would be particularly useful during the hours of peak usage, such as the beginning or the ending of the physician shift rounds (Siegel, Protopapas et al. 1997). The tools available for diagnosis should adapt to the utilization pattern corresponding to the user’s role, and to the type of disease investigated as reflected by the information supplied by RIS/HIS (Horii, Feingold et al. 1996). The fact that a study found no statistical difference in the outcome of interpretations between a PACS diagnostic workstation and a cheaper, PC-based review workstation, encourages building a more extensive network of workstations at the point of clinical care delivery (Parasyn, Hanson et al. 1998).
In some clinical environments, a money saving strategy could allow the ICU physicians to perform the first interpretation of radiology examinations (Simon, Khan et al. 1996). This is not a new trend, since ICU physicians often act as substitutes for other specialists, such as cardiologists, gastroenterologists, orthopaedists and general surgeons. For better patient care, radiologists should still be consulted on high risk patients or on exams difficult to interpret. Therefore PACS should support preliminary interpretation from ICU physicians; this preliminary interpretation should be available when the radiologist dictates the final report (Siegel, Groleau et al. 1998).

3.10.3 Reliable and fast -- new toolkit and user interface

Reliable (up-time 99.999%) and fast -- these are the most important objectives for radiology workstations located in the ICU. The use of two or four monitors is recommended to reduce the time for image review and interpretation when compared with one monitor workstations (Reiner, Siegel et al. 1997).

The user interface should be tailored to the main task facing the ICU staff. If the task consists of image review, the tools used for radiology interpretation should be made available. If the task consists of image interpretation, extra attention and possibly more advanced image manipulation tools may be required. Whenever the study displayed has a radiology report associated, a reduced, disease oriented set of features for ‘streamlined’, ‘heads-up’ reading should be presented to the user (Horii, Kundel et al. 1994) (Horii, Feingold et al. 1996). Some tools, such as W/L, zoom and pan, should be ‘always on’ through a dedicated mouse button. Due to wide variations in radiographic exposure, the automatic adjustment of brightness and contrast can significantly reduce the W/L manipulation required for historical study comparison (Rottenberg, Chin et al. 1996). CR display presets optimized for specific findings and pathologies could facilitate image comparison and standardize display across examination type (Andriole, Gould et al. 1999). Another criterion relevant for the user interface customization is the manipulation pattern associated with the type of study: for paediatric ICU, W/L and invert will be often used to detect the tips of the catheters or edges on pulmonary opacities, while for neonatal ICU, zoom is used for detection of subtle tubes (Brill, Winchester et al. 1996)
3.10.4 Automation of most routine tasks

Productivity will be high when the interpretation time, user fatigue and the tools usage is reduced without negative impact on patient care. An ICU oriented hanging protocol should deal with automatic loading of relevant priors, automatic image orientation and automatic shutter¹, especially for paediatric images (Evanoff and McNeill 1997). In order to encourage the use of image processing tools by the ICU personnel, a ‘return to original’ and one level ‘undo’ should be implemented and easily accessible.


3.11 Workstation evaluation

When evaluating radiology workstations, researchers looked into several criteria, such as workstation requirements, functionality, and usability (Kolodny, Raptopolous et al. 1999; Erickson, Ryan et al. 2001; Moise and Atkins 2002a). Observations of the radiologist workflow revealed three requirement categories: user control of image management, navigation of images and image series, and simultaneous availability of detail and context (Seltzer, Judy et al. 1994; van der Heyden, Inkpen et al. 2001). Several papers evaluated different methods for displaying medical images, with respect to the screen real-estate problem: how to provide enough detail to allow for diagnostic interpretation, while preserving the contextual information (Beard 1990; Seltzer, Judy et al. 1994; van der Heyden, Inkpen et al. 1999a; van der Heyden, Inkpen et al. 1999b; van der Heyden, Inkpen et al. 2001). The results showed no significant difference between the ‘thumbnail’ technique and the ‘detail-in-context’ technique. However, these studies did not include stack mode viewing, which is more suitable than tile mode for reading cross-sectional imaging, as shown in Section 3.2.

---

¹ Removal of unexposed background, which is displayed as bright white on screen.
3.12 Usability

Usability of a computer system can be defined as the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component (IEEE 1990). Usability testing refers to the evaluation of information systems that involves participants who are representative of the target user population (Kushniruk, Patel et al. 1997). The radiology workstation design should balance the addition of features with the cognitive workload on the radiologist to manipulate these tools, and with the brain fatigue this secondary process is introducing. Failure to do so will result in reduced productivity, greater training requirements, incomplete utilization of the system’s features, and slower adoption rates.

Conventional methods of evaluation, such as questionnaires and interviews with users, are limited by the user’s recall of their experience in using a computer system. The feedback provided by these methods only refers to what the users think they do while using the computer system, which can be considerably different from the actual users’ behaviour. Outcome-based evaluations focus on examining effects of system use on clearly defined and pre-specified outcome measures, but do not allow the examination of the actual process of the system in day-to-day activities. Cognitive task analysis is concerned with characterizing the decision making and reasoning skills of subjects, as they perform activities involving the processing of complex information (Kushniruk, Patel et al. 1997). Subjects may be asked to ‘think aloud’ as they perform specific tasks, or asked to write descriptions of the tasks they are performing. Subjects can also be video-recorded in order to capture the physical, temporal and social context as they interact with the system tested. Contextual inquiry involves going to the user’s actual work environment and interviewing and observing the users while they perform their work (Coble, Maffitt et al. 1995).

Often times radiologists receive little or no training from the PACS vendor on how to use a complex radiology workstation system. Consequently, the ‘threshold’, or how difficult it is to learn how to use the system, is an important consideration for designing successful radiology workstations. One study looked into how easy it is to learn to use a particular radiology workstation, and how intuitive the controls are.
(Erickson and Kossak 2000). When rating a radiology workstation, it is also important to take into consideration the workstation’s ‘ceiling’, or how much can be done using the system. For example, how many exams can be reviewed in a day by a radiologist that is also an expert user of the workstation. Unfortunately, the study reported above did not provide information about the workstation’s ceiling.

### 3.13 Ergonomic user interfaces for the radiology workstation

A poor, uncontrolled user interface (UI) design leads to systems that are difficult to use. The user could spend more time dealing with the interface than the actual task the application is supposed to support. Because film interpretation was optimized through years of training, experience, and incremental improvement of film-handling processes, the film workflow provides a robust, versatile exception handling, and complex manual system. For softcopy interpretation to offer the same level of productivity, the user interface must be carefully designed to prevent any overhead required for manipulating medical images in a digital environment, overhead that does not exist when reading film from traditional lightboxes. One important goal of cognition-based ergonomic user interface design for radiology workstations should be to reduce the number of fixation clusters on the area of the menu used to control display functions, which was reported to be 20% of all screen fixations (Krupinski and Lund 1997). We should also focus on transferring most of the user-workstation interaction to the subconscious level of the human cognitive skills, thus reducing the interruption in the task of diagnosis.

### 3.14 Providing a consistent metaphor

User centric design is one condition for well-integrated workstation functionality (Moise and Atkins 2002a). The application should provide a consistent paradigm, by building a mental model for the user that should naturally integrate the workstation functionality (Arenson, Chakraborty et al. 1990). A unity should be visible between the workstation implementation, the operating system in use, and the UI elements of the workstation, such as colour scheme, menus, and dialog boxes. The success of implementing this goal should be assessed using metrics like ergonomics and user fatigue.
Good UI design is often based on metaphor. Invented in the early-1970s at Xerox's Palo Alto Research Center, the 'desktop metaphor' is a well known metaphor, and represents the term frequently used for the hierarchical system of files, folders and icons that we use to manage the information from our computers, which used to be stored on the top of our desk.

Some metaphors used in radiology workstations design include:

- The ‘photo album’: in order to retrieve medical images, a book-like interface element contains all images for a specific modality and/or anatomical region.
- The ‘thumbnails’ (van der Heyden, Inkpen et al. 1999b; van der Heyden, Inkpen et al. 2001) or the ‘digital light projector’ (Demiris and P 1997): these small icons are used to give an all-round impression of the data set, allowing for the orientation and navigation in an image volume.
- The ‘tile mode’ and ‘stack mode’, presented in more details in Section 3.2.
- Zooming: this feature is equivalent to the radiologist getting closer or further away from the film, allowing the radiologist to zoom in and out of the selected image.

### 3.15 Hanging Protocols

#### 3.15.1 Problem statement

In the film-based environment, some of the most time consuming tasks from the interpretation workflow include:

- Retrieval of prior examinations (from film archive) and associated radiology reports (from RIS), which is generally done by file room clerks.
- Hanging the images for interpretation, which is generally done by technologists.
- Spelling and formatting of reports, which is generally done by transcriptionists.

It is often the case that switching from hardcopy to softcopy interpretation involves more responsibilities for the radiologist, as he or she is now responsible for the tasks described above. Additionally in many institutions radiologists are remunerated on the number of studies on which they report, so any technology or workflow alteration which requires extra time for the radiologist is bitterly resented. A successful transition from film-based reading to softcopy reading is conditioned, to a large extent, by
replacing many of the functions done manually with automated equivalents, such as hanging protocols.

Hanging protocols (HP) were created to alleviate the increased workload on the radiologist, by automating image presentation on computer screens immediately after opening a case (Strickland and Allison 1995). By automatically displaying images on multiple monitors the requirement for users to interactively manipulate images is minimized. HPs usually include information regarding placement of the sequences, viewing mode, layout, W/L settings, zoom and pan (Moise and Atkins 2002c).

3.15.2 Hanging protocols: Definition

In a film-based environment, a technician manually hangs the film examinations on the lightbox in a known examination-specific default, prior to examination by the radiologist (Blume, Bergstrom et al. 1998). For viewing large numbers of films, an ‘alternator’ can be used to display up to 8 films at a time, as shown in Figure 1.

For softcopy interpretation the technician’s work is replaced by hanging protocols which are meant to partially reduce or completely eliminate the time spent by radiologists in rearranging images on the monitors. By hanging protocols (HPs) we mean a set of rules defining how images are arranged on one or more computer monitors immediately after opening a case. The automatic image presentation streamlines the radiologist interpretation task.

Hanging protocols provide potential for faster reading and improved diagnostic confidence. HPs can contribute to the development of guidelines for softcopy reading of radiographic studies and practices for optimal readout of specific cases. Standardization of the description of these protocols would allow institutions to share protocols for comparison and evaluation.

3.15.3 Related relevant work

Folder manager concepts, such as prefetching, auto-routing, storage and image presentation, describe how appropriate data are delivered where and when (Bellon, Feron et al. 1996b; Badano and Flynn 1997; Nahmias, Kenyon et al. 1997; Blume, Bergstrom et
al. 1998; Patel and Kushniruk 1998; Mendenhall, Dewey et al. 2001). In the film-based radiology department these functions are carried out by different clinical personnel, but in PACS all these functions need to be automatically performed by computers (see Table 2 for details).

Table 2: Folder manager comparison: hardcopy vs. softcopy

<table>
<thead>
<tr>
<th>Folder manager concept</th>
<th>Definition</th>
<th>Hardcopy</th>
<th>Softcopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefetching and auto-routing</td>
<td>Retrieval of relevant prior studies and delivery of data to anticipated locations of need</td>
<td>Film file room clerk</td>
<td>HPs and PACS agents interacting with RIS / HIS</td>
</tr>
<tr>
<td>Storage management</td>
<td></td>
<td>Film filing by x-ray technologist</td>
<td>Hierarchical storage</td>
</tr>
<tr>
<td>Image presentation</td>
<td>Presentation of data for interpretation</td>
<td></td>
<td>HPs</td>
</tr>
</tbody>
</table>

The idea of controlling the initial display of images is not new. In 1990 “automatic schemes” were introduced (Arenson, Chakraborty et al. 1990), while a 1996 paper describes a “montage-display format” of 20 selected images with significant features or pathology (Lou, Huang et al. 1996). Strickland et al. presented default modes for the display of images (Strickland and Allison 1995; Strickland, Allison et al. 1997). A group from University of California at Los Angeles introduced Structured Display Protocols that model the presentation of data according to the diagnostic task, represented using specific Unified Modelling Language™ diagrams (Valentino, Wei et al. 2001). For example, the Digital ViewBox follows the workflow of brain tumour and spine interpretation using “protocols, modes, and tools,” where a protocol has modes that represent diagnostic reading tasks (Harreld, Valentino et al. 1998; Valentino, Harreld et al. 1998). Structured Display Protocols failed because the protocols were too entwined with the user interface.
In order to acknowledge the importance of the hanging protocols, DICOM Working Group Eleven started in 2000 to work on standardization of the representation of the Unified HP model.

3.15.4 Current limitations and challenges associated with HP

The early implementations of HP were primitive, lacking user control over series placement, W/L settings, as well as the zoom factor and panning required. Even recent implementations of HPs fail to describe the scrolling interaction, the underlying layout algorithm, complex screen layouts that cannot be represented as rows by columns, the series linking for synchronized navigation, or the display of flagged images only.

In conclusion, the two major drawbacks with current HPs are:

- they provide only the initial presentation of images. This initial presentation may be used solely to provide a gestalt on the study to be read.
- they do not handle other clinical information normally reviewed by radiologists, such as prior radiology reports and the patient’s history.

Other challenges and limitations associated with current HPs are presented below.

3.15.4.1 Series placement

When a HP is created, the position on the screen of each series is recorded as series identifier. This identifier comes from the ‘ID series description’ field present in the header of each DICOM image. For HPs to work, consistent naming mechanisms for the series should be used. When series naming is not consistent, series identification and placement is compromised.

Unfortunately, series names are sometimes missing - most CR chest images do not include identifiers in the DICOM header, such as ‘PA’ for posterior-anterior projections, and ‘LAT’ for lateral projections. One solution for this problem relies on the consistent ordering of the series by the technicians, which involves more training. For example, in a CR examination, the first series is always the ‘PA’; if a second series exists, that will be the ‘LAT’. Another solution consists of image processing algorithms for detection of ‘PA’ and ‘LAT’ series.
Series names are sometimes inconsistent because different technicians are in charge of the acquisition of medical images (Guld, Kohnen et al. 2002). Different technicians may use different names to describe the same series, such as ‘Chest’ and ‘CR/Chest’. Typing errors are also responsible for the different series naming. This type of situation could be corrected by eliminating manual, error prone input through the use of DICOM worklists, combined with predefined acquisition protocols with consistent series labelling for modalities. Technicians will choose a scanning protocol, for example ‘MR brain with contrast’, with a predefined format, such as first series ‘T1 axial’, second series ‘T2 axial’, and third series ‘Contrast’. Using worklists will also decrease the technologists’ workload by eliminating the manual steps required for setting up the parameters (such as the series names) of the scanning protocols.

3.15.4.2 Layout

A flexible workstation design should allow for a hierarchical layout configuration. At the root level, the display area should be divided into several ‘virtual monitors’. A virtual monitor is usually mapped onto a physical monitor (Ghosh, Andriole et al. 1998). If the design for the radiology workstation assumes ‘one study per monitor’, it is imperative to be able to assign two virtual monitors on a physical monitor. This will allow the user to compare two or more studies for the same patient when using workstations with only one physical monitor.

The second level of layout control should be at the level of each individual monitor, allowing users to independently control the number of ‘viewports’ on each monitor. For example, the user could set up one viewport for a single CR image on the first monitor, and 4 viewports or a 2x2 layout for an MR study on the second monitor.

3.15.4.3 Brightness and contrast settings

Images are generally displayed for soft copy presentation on diagnostic greyscale monitors. These monitors provide high resolution, good light output, and luminance uniformity correction. These monitors can also be calibrated for image conformance and consistency. The radiology workstation is responsible for mapping pixels intensity values stored in the DICOM images, or up to 65536 levels of grey for images acquired at 16 bit
pixel, onto monitors with less than 256 shades of grey. This is usually done by applying a linear transfer function like W/L, with its variants: min/max, and slope/intercept. For non-linear transfer functions, a DICOM Modality, Presentation or volume of interest look-up table can be used, but they are rarely used in practice. For the scope of this thesis, only the W/L solution is considered for adjusting for brightness and contrast.

It is important to be able to customize the default W/L values when loading series on screen. It may also be useful to replicate series for display with different contrast settings. For example, a CT abdominal scan can be presented with both 'Lung' and 'Liver' W/L values, as shown in Figure 7 and Figure 8, respectively. Both of these options can considerably reduce the time required by the radiologists to CT or MRI images with the optimum brightness and contrast values.

3.15.4.4 Loading of prior examinations

If prior studies are stored on slow archive devices, these studies should be retrieved in advance using prefetching algorithms. The likelihood of obtaining a prior examination that matches the current one was investigated with selection of priors based on their number (2 priors 83%, 4 priors 91%) or defined time interval (one month for 80%). More complex strategies based on metagroup categories for examination type mnemonics resulted in a sensitivity of 98.3% and a specificity of 100% (Andriole, Avrin et al. 2000). These algorithms should be coupled with automatic loading of the next examination. By loading images for the next study in the background (before the user closes the current study), the delay for loading the next study is completely eliminated.

The complexity of arranging studies on-screen can increase dramatically when prior examinations must be automatically loaded. Selection of significant prior examinations should take into consideration relevant modalities, procedure types, body parts, availability of the report, the date of the study, the type of media the study is stored on, and the hardware workstation configuration.
4 ‘Historical evolution’ of our line of work

4.1 Hands-on industrial experience and direct interaction with the radiologists

My interest in radiology workstation design, medical image display, and automation of radiologists’ reading steps started in 1998, when the author became involved in the design and development of a multi-modality diagnostic reading workstation for A.L.I. Technologies, Richmond, Canada. During the 3.5 years of research and development with A.L.I. Technologies, the author had an extensive one-to-one interaction with radiologists from various locations in the United States, which helped me gain domain expertise, and understand the radiologist’s perspective on softcopy interpretation. For this thesis, we also took advantage on my practical expertise on developing a state-of-the-art radiology workstation, which provided insight into the vendor’s perspective to softcopy interpretation.

4.2 Medical imaging conferences and trades shows

We have attended some of the most important medical imaging conferences and trade shows, such as The International Society for Optical Engineering (SPIE) Medical Imaging conferences and the Radiology Society of North America (RSNA) Scientific Assembly and Annual Meeting. RSNA is arguably the premier radiologic meeting of the year, highlighting the best in science, technology and education. The author had the opportunity to communicate with many front-line decision-makers and medical imaging
leaders from all over the world. The author discussed his research objectives with radiology professionals who are also involved in state-of-the-art technologies, thus helping to advance both the art and science of quality softcopy reading.

The author also attended the annual scientific and educational meeting presented by the Society for Computer Applications in Radiology (SCAR), a leading educational forum on diagnostic imaging and information interpretation/management. This symposium features the most recent developments in medical computer applications, particularly advances in computer technology that improve the clinical practice of radiology and the effective management of health care resources. SCAR showcases technical exhibits of leading providers of radiology computer and information technology hardware, software, and services.

4.3 First user study: radiology workstation design

We performed a questionnaire-based user study in November 1999 in an outpatient radiology clinic equipped with state-of-the-art PACS and imaging scanners. This research has been approved by the Simon Fraser University Office of Research Ethics (see Appendix H: Approvals from the Office of Research Ethics). This study took place in Beverly Hills, California, and involved eight radiologists who answered questions and provided comments on three complementary research topics described below. The first section of the study assessed the benefits of softcopy vs. hardcopy, and it was meant to provide feedback for improving the design of radiology workstations. The second section of the survey obtained radiologist feedback on the use of hanging protocols (Moise and Atkins 2002c). In the third section we investigated the impact of display devices on the radiologists’ workflow. The questionnaire we used for this user study is presented in First user study questionnaire.

4.3.1 User profiles

All eight radiologists involved in our study were familiar with softcopy interpretation. Prior to our study, each radiologist used at least 3 different radiology workstations, and was involved in softcopy diagnostic interpretation for at least 3 years. Two radiologists specialized in cross-sectional examinations, two radiologists focused on
projection radiography and the other 4 radiologists specialized in US radiography. Table 3 summarizes the average daily workload (number of cases read per day) distribution for these radiologists (based on their own estimation) in either hardcopy or softcopy environment. RF stands for Radio-fluoroscopy.

### Table 3: Profiles of the radiologists involved in our study

<table>
<thead>
<tr>
<th>Rad.</th>
<th>Num. of rad. wrkst. used</th>
<th>Years of softcopy reading</th>
<th>Number of studies read per day</th>
<th>Hardcopy</th>
<th>Softcopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT/MR</td>
<td>CR/RF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT/MR</td>
<td>CR/RF</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>5</td>
<td>18</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3 1/2</td>
<td>30</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>4</td>
<td>7-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>50-60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>8</td>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The goal of this study was to determine the requirements for designing third generation radiology workstations. We asked the radiologists to provide the pros and cons for both hardcopy and softcopy interpretation reading. It was clear to us that for the softcopy paradigm to succeed, we had to avoid a one-to-one translation of the steps involved in the hardcopy-specific workflow. Workflow reengineering associated with the use of PACS has resulted in increased efficiencies of the technologists by 20–60%, of the clerical staff by more than 50%, and of the radiologists by more than 40% (Siegel and Reiner 2002a).
4.3.2 Softcopy vs. hardcopy

Table 4 summarizes the drawbacks of hardcopy reading. The most commonly mentioned drawback of hardcopy interpretation, as expressed by 7 of the 8 users, was the inability to perform post processing operations, such as changing the W/L settings. Other reported disadvantages for the use of hardcopy were difficulty in handling increasingly more images, inability to conveniently zoom, measure and annotate, and difficulty in locating films from previous examinations.

Table 4: Drawbacks of the hardcopy interpretation mode

<table>
<thead>
<tr>
<th>Rad.</th>
<th>MR/CT</th>
<th>CR/RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No cross reference; too many films for view box capacity; availability of old study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Post processing missing (W/L, zoom)</td>
<td>Inconvenience to see previous exams; lost film or signed to doctors</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Too many films; cannot post process measurements</td>
<td>Too many films</td>
<td>It cannot enlarge</td>
</tr>
<tr>
<td>4</td>
<td>Stack missing, post processing (W/L, Hounsfield measurement)</td>
<td>W/L, zoom</td>
<td>Comparison available; more efficient on computer</td>
</tr>
<tr>
<td>5</td>
<td>No W/L changes, finding and location of the previous films if applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Post processing (magnifying glass); increased number of images for bone/soft, W/L settings; getting previous exams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Exposure, quality, lost films</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>It cannot change W/L (post processing); access to previous studies; it can be lost</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 summarizes the drawbacks reported by radiologists about the softcopy interpretation process, reflecting design flaws typical to second generation radiology
workstations. Transfer speed was slow due to limitations in file transfer protocols used, such as the DICOM standard for image communication. The lack of usable hanging protocols aggravated this situation by increasing the amount of time wasted by radiologists from study selection until all images were displayed in a format suitable for interpretation.

Table 5: Drawbacks of the softcopy interpretation mode

<table>
<thead>
<tr>
<th>Rad.</th>
<th>MR/CT</th>
<th>CR/RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stability, lack of detail in context,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>compatibility between systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Speed of displaying; comparing (space</td>
<td>Setup desktop (window position, size not</td>
<td>UI not</td>
</tr>
<tr>
<td></td>
<td>limitation)</td>
<td>not fix)</td>
<td>friendly (SUN)</td>
</tr>
<tr>
<td>3</td>
<td>Time spent, setting up viewing; too many</td>
<td>Resolution, slow speed, complexity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Comparing studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Time</td>
<td>W/L, Cine play</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Resolution; loading and arranging time</td>
<td>Learning curve, poor printing</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Despite all the drawbacks listed above, the radiologists involved in our study showed strong support for softcopy interpretation. Some of them even mentioned they will “never go back to reading film.” This ‘enthusiasm’ is very likely related to the set of features commonly used by radiologists for softcopy interpretation, as shown in Table 6. Image manipulation tools can potentially increase the amount of information relevant to the diagnosis process, thus increasing the accuracy and confidence of the radiological interpretation. Since inability to adjust W/L settings was the major drawback reported for hardcopy interpretation, it comes at no surprise that W/L processing was the most commonly used tool for softcopy interpretation, together with image zoom and pan. For cross-sectional imaging, ability to link (navigate synchronously) multiple series
displayed in stack mode was mentioned by most of the radiologists involved in CT/MR interpretation.

Table 6: Most used features for the radiology workstation as of December 1999

<table>
<thead>
<tr>
<th>Rad.</th>
<th>MR/CT</th>
<th>CR/RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>W/L, cross reference, zoom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>W/L, zoom, clip and zoom, measure, HP, linking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>W/L, zoom, clip and zoom, measure, HP, linking</td>
<td>Invert/zoom/pan</td>
<td>Comparison</td>
</tr>
<tr>
<td>4</td>
<td>Link, measurements, region of interest, W/L</td>
<td>W/L, and zoom</td>
<td>Less processing and measurements done by tech.</td>
</tr>
<tr>
<td>5</td>
<td>Zoom, magnifying glass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>W/L, measure, link (more for MR than CT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>W/L, zoom, measure</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>W/L, zoom, invert, measure, magnifying glass (edge enhancement - compensates for loss in resolution)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7 summarizes the ‘wish list’ for the radiologists involved in our study. Cross-sectional specific features include advanced mark-and-measure, such as labeling the vertebrae or measuring the Hounsfield units (H.U.) on CT studies, multi-planar reformattting (MPR) and inter-study linking. All these features were demonstrated during RSNA 2001. The wish list also includes rotation with an arbitrary angle, open-study notification (to signal when a study is opened for interpretation by another radiologist), improved viewing protocols and ability to save the user’s arrangement of images for interpretation.
Table 7: Most useful new features required for the radiology workstation as of December 1999

<table>
<thead>
<tr>
<th>Rad.</th>
<th>MR/CT</th>
<th>CR/RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MPR; linking two studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HP- list of freq. used protocols; enhanced mark and measure capabilities (numbering on spine)</td>
<td>Know if somebody is also looking at the study; saving user settings when doing a study</td>
<td>Know if somebody is also looking at the study; saving user settings when doing a study</td>
</tr>
<tr>
<td>3</td>
<td>Healthy study (H.U. measurement continuous state)</td>
<td></td>
<td>Zoom, marks and measurements</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Display healthy study example</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Linking images from two different exams; calibrated images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Edge enhancement</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Viewing protocols; fast and reliable; arbitrary rotation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3.3 Hierarchical HP

In this section, our goal was three fold: to assess the need for developing HPs, to validate our hypothesis towards a hierarchical structure of HPs, and to determine the average number of HPs required, so we can estimate how practical it will be for radiologists to define and use HPs.

We investigated the perceived impact on the radiologist’s productivity and accuracy due to switching from hardcopy to softcopy presentation. Table 8 summarizes each radiologist’s estimation of the average time spent on ‘image organization’ (i.e. the time spent preparing a study for interpretation) and on the quality of the diagnosis. The quality of the diagnosis hardcopy vs. softcopy was rated:
- ‘>’, as in better for hardcopy,
- ‘=’, as in about the same
- ‘<’, as in better for softcopy.

These results stressed an increasing acceptance for softcopy reading and the importance of automatic image organization through HP.

Table 8: Hardcopy vs. softcopy – quality of diagnosis and image organization time

<table>
<thead>
<tr>
<th>Rad.</th>
<th>Quality of diagnosis</th>
<th>Image organization (min)/Accuracy hardcopy vs. softcopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hardcopy vs. Softcopy</td>
<td>Hardcopy</td>
</tr>
<tr>
<td>1</td>
<td>&lt;</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>3</td>
<td>=</td>
<td>&gt;/&lt;</td>
</tr>
<tr>
<td>4</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>5</td>
<td>&lt;=</td>
<td>&lt;</td>
</tr>
<tr>
<td>6</td>
<td>&lt;</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>&lt;</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>&lt; (lumbar)</td>
<td>=(chest)</td>
</tr>
</tbody>
</table>

The HP selection paradigm that we evaluated consisted of the following filters: modality, procedure type (see ‘aliases’), body part (anatomy), name and number of series, number and type of relevant prior examinations, and configuration of the workstation displays. All the radiologists were interested in our idea of organizing the HPs in a hierarchical structure: Root->Modality->Body part->Procedure type->Priors. A matching HP can be searched on the deepest level on the tree. If the node is empty the search should go up in the tree, eventually reaching the modality level. However, this hierarchical HP organization does not offer a perfect solution for choosing the order.
(priority) of the filtering. The HP tree presented for their verbal validation is illustrated in Figure 31.

<table>
<thead>
<tr>
<th>General</th>
<th>MR</th>
<th>Extremity</th>
<th>Knee</th>
<th>LeftKnee</th>
<th>LeftKneeWithContrast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RightKnee</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arm</td>
<td>LeftArm</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spine</td>
<td>SpineLumbar</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SpineCervical</td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td>Chest</td>
<td></td>
<td>ChestLatera</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extremity</td>
<td></td>
<td>LowerExtremity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UpperExtremity</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 31: Hierarchical HP organization*
We inquired about the number of HPs that each radiologist would use to cover most of the daily work. We expected the number of the HPs to be proportional to the number of body parts imaged with each modality. Table 9 shows the estimated number of HPs each radiologist thought would be required, and the percentage of the studies that will be covered by these HPs.

**Table 9: Estimated number of HP and percentage of the studies covered.**

<table>
<thead>
<tr>
<th>Rad.</th>
<th>Estimated # of HP req.</th>
<th>Percentage of studies covered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CT/MR</td>
<td>CR/RF</td>
</tr>
<tr>
<td>1</td>
<td>&gt;15</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3/10</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>15/5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>50-100</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>15-20</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We also investigated the impact of display devices on the radiologist’s workflow. The radiologists were familiar with monitors with different properties: grayscale and color, portrait and landscape, and spatial resolution of 1, 2, 3 or 5 mega-pixels. We investigated the importance of the spatial resolution by comparing side-by-side 2 types of high-resolution grayscale monitors: 1728x2304 vs. 2048x2560. We used a 0 to 4 scale, with 0-not important, and 4-most important. We inquired about the ‘frame effect’ produced by 2 adjacent monitors: ‘Yes’ means the radiologist would prefer a bigger monitor to replace the 2 smaller ones, ‘No’ means the frame effect is not disturbing. The preferred number of monitors (1-4) and the preference for a general, multimodality workstation (‘Yes’) vs. a suite of specialized workstation (‘No’) are also presented in Table 10.
Table 10: Effect of display devices: general vs. specialized workstation; is frame effect disturbing? preferred number of monitors, and importance of spatial resolution.

<table>
<thead>
<tr>
<th>Rad</th>
<th>General vs. spec.</th>
<th>Frame effect</th>
<th>Preferred number of monitors</th>
<th>Importance of spatial resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT/MR</td>
<td>CR/RF</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>Yes</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

4.3.3.1 Discussion

Our hypothesis on the radiologist’s need for an automatic arrangement of images was confirmed. Based on conversations with the radiologists, and from the results presented in Table 8, HPs have the potential to reduce the interpretation time by 10-20%. The data collected suggests that 10-20 HPs / modality can accommodate the radiologist’s needs in 80-90% of the studies displayed for interpretation. Less than twenty HPs per modality will not be too difficult for radiologists to generate, nor for radiology workstations to manage and select from. The remaining 10-20% of examinations will be impractical to cover under HPs. These remaining studies represent either very rare examination types or exceptions, that will occur in situations such as when the patient moved and the technician had to add another series, or when the study was sent twice from modality to PACS so each series was duplicated.

These data increased our confidence the HPs will be critical in combating the limitations of the screen real-estate, especially for complex situations like MR
brain/angiography or ICU when comparison with prior examinations becomes critical (Moise and Atkins 2002b).

The hierarchical structure we proposed in November 1999 was very well received at that time. However, further refinements could contribute further to an improved softcopy interpretation, as we show in the following sections.

4.4 Softcopy interpretation: a literature survey

Since 2001 we surveyed the literature on different topics related to PACS and softcopy interpretation. We were interested in the evolution of radiological interpretation, and the factors which accelerated the transition from hardcopy to softcopy reading, such as technological advances and increasingly more digital modalities, as described in Section 2.1.1. We also read on the advantages and disadvantages associated with both hardcopy and softcopy reading, as summarized in Sections 2.1.3 - 2.1.5. Other topics surveyed include workstation requirements, design, and evaluation, the perceptual and cognitive aspects related to lesion detection, accuracy of interpretation, visual search and satisfaction of search.

4.5 Determining the requirements of diagnostic workstations

There are several approaches for determining the requirements of diagnostic workstations (Horii 2002):

- Observing and then interviewing radiologists
- Building functional models of radiologists’ working tasks and combining these with principles of human-computer interface design
- Using an interactive process consisting of building prototypes, evaluation by radiologists, and prototype refining. This process would stop when the radiologists express a high level of satisfaction with the result.

In the following section we present our inter-disciplinary approach to identify key components of the radiologist’s task, by using well established theories from different research domains. By surveying the psychology literature, we identified research results from perception and cognition applicable to radiology softcopy reading. By studying
perception we gained insight on the main causes for false negative errors. We used cognitive research to draw on the knowledge of structures and cognitive processes underlying radiologist performance.

4.6 Perception

A unifying principle in the psychology of visual perception has been that unless the perceiver makes assumptions about the physical world which gave rise to a particular retinal image, perception just isn’t possible (Bruce and Green 1985).

Radiology training is designed to provide assumptions about normal and abnormal medical images. However, radiologists do make mistakes. General estimates suggest that, overall, there is a 20-30% miss or false negative rate in radiology, with a 2-15% false positive rate, a 10-20% inter-observer variation and 5-10% intra-observer variation (Hendee 2002). False positives can be explained by image noise and overlapping anatomic structures that often mimic disease entities. However, false negative errors are harder to understand, especially when missed lesions can be seen in retrospect (Krupinski 2000). Based on gaze dwell time, three classifications have been found for false-negative errors: faulty visual search or failure to fixate the abnormal region, faulty pattern recognition or failure to report a lesion that has been fixated for a short amount of time, say less than 1 second, and faulty decision making or failure to report a lesion that has been extensively fixated, say more than 1 second.

One study suggested a measurable conspicuity index for the physical characteristics of naturally occurring pulmonary nodules. This index could be used for synthesizing artificial nodules (Manning, Ethell et al. 2002). The mean values of the synthesized lesions were higher than the natural lesions, but did not result in a higher detection rate, and the observers were unable to distinguish real nodules from artificial ones in a direct test of their ability to do so. Nodules with an average conspicuity were detected and recognized with greater confidence than low or high conspicuity nodules. This led to the conclusion there may be a perceptually preferred range of conspicuity for signals in complex fields. The location of a nodule may have an independent influence on its detection likelihood, but the eye tracking data did not support this hypothesis,
indicating an effective search strategy on the part of expert observers. Observer interpretation of multiple abnormality images frequently resulted in the termination of search prior to verbal acknowledgement of all abnormalities. However, analysis of eye tracking sequences revealed that the majority of missed lesions within the nested series were fixated. The authors’ interpretation of this finding was that the observers were making recognition errors on occasions, and although they sometimes detected a nodule they misinterpreted its significance.

Perception research will benefit clinical radiology by providing methods for reducing observer error, providing objective standards for image quality, and providing a scientific basis for image-technology evaluation. The Medical Image Perception Society was created in 1997 to promote research in (Krupinski, Kundel et al. 1998):

- mathematical modelling of the detection of discrete abnormalities in noise-limited images,
- understanding how observers find discrete abnormalities in images when their locations are unknown,
- understanding how knowledge and experience affect the recognition and detection of abnormalities,
- developing perceptually based standards for image quality,
- developing computer-aided perception tools, and

4.6.1 **Understanding the visual search**

The noise in medical images limits the perception of contrast and detail. Current research is extending the mathematical models that relate signal detection to physical descriptions of image signal and noise to include realistic lesions and anatomical backgrounds.

Many lesions can be seen by peripheral vision and verified by foveal vision, but peripherally inconspicuous targets must be found by scanning the fovea over the image. The resultant search pattern, as shown by eye-position recording, is influenced by both the patient’s clinical history and the radiologist’s experience (Kundel, Nodine et al. 1978).
While radiologists' search patterns on film and monitor are comparable on diagnostic image areas, scanning non-image areas differs significantly (Krupinski and Lund 1997). Radiologists generally scan to menu within first 10 sec of search, after only about 3 fixations on the diagnostic image, as shown in Table 11. A possible explanation of this behaviour is the poor initial image presentation, which requires adjustment using the image processing menu. Since radiologists scan to the menu extremely early in the search, this may cause a disruption in initial image perception, which may delay the detection of lesions compared with film.

Table 11: Mean percentages of image areas receiving fixation clusters during search for CRT monitor vs. film viewing of skeletal radiographs. From (Krupinski and Lund 1997).

<table>
<thead>
<tr>
<th>Image Area</th>
<th>Film</th>
<th>Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Image</td>
<td>95</td>
<td>78</td>
</tr>
<tr>
<td>Background</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Image Processing Menu</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

When compared to hardcopy, softcopy viewing takes about 60 seconds longer in skeletal computed radiography images. The time to first fixate the lesion is significantly longer on CRT (4.67 s) than film (2.35 s), and readers generated more clusters (31.98 vs. 27.55; p < .05) with longer dwell times (892 ms vs. 816 ms; p < 0.05) on diagnostic areas of images on the monitor than on film images, as shown in Table 12.
Table 12: Median dwell times (ms) for decisions made during search of skeletal images for trauma on film vs. a CRT monitor (6 readers). Only true-negatives differ significantly. From (Krupinski and Lund 1997).

<table>
<thead>
<tr>
<th>Decision</th>
<th>CR Film</th>
<th>CRT Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>True-Positive</td>
<td>1206</td>
<td>1286</td>
</tr>
<tr>
<td>False-Negative</td>
<td>1204</td>
<td>938</td>
</tr>
<tr>
<td>False-Positive</td>
<td>896</td>
<td>895</td>
</tr>
<tr>
<td>True-Negative</td>
<td>358</td>
<td>532</td>
</tr>
</tbody>
</table>

### 4.6.2 Satisfaction of search

Satisfaction of search (SOS) refers to the fact that detection of one radiographic abnormality may interfere with detection of other abnormalities on the same examination. In other words, in viewing radiographs there is a tendency to become satisfied after identifying the first abnormality, which may lead to failure to search for additional findings. This is a tendency germane to the interpretation of all forms of imaging, and a likely source of diagnostic errors. In one experiment, the observers read 15 cases with single findings and 15 cases with multiple findings (Ashman, Yu et al. 2000). Out of 15 cases, the average detection rate was 11.25 and 11.72 for one finding among cases with single and multiple findings, respectively. However, the detection rate for the second and third finding (among the cases with multiple findings) dropped down to 6.12.

There are several theories that try to explain SOS. The theory of strategic termination of search is based on the minimization of the number of false-positive diagnoses. The adherence to a perceptual set promotes the inclination to discount findings of a different category from the ones found first. Another theory hypothesized a lack of attention to regions that did not contain contrast material on contrast-enhanced images, or that visual distractors, such as bright blood vessels, impaired the observer's ability to detect bright liver nodules on contrast enhanced spiral CT scans (Wester, Judy et al. 1997).
The traditional theories for SOS errors, that the search is terminated after the discovery of an abnormality, have been discredited, after it was shown that observers continue to inspect images after an initial abnormality was reported. According to the Kundel-Nodine method of error classification, an unreported abnormality is assumed to have been recognized if the gaze time of the abnormality exceeds some estimated value (Kundel, Nodine et al. 1978). This theory is challenged by the fact the eyes can fall on an abnormality and the visual pattern may be analyzed as not corresponding to the abnormality. An alternate theory, protocol analysis, borrows from research on human problem solving: while attempting to solve a problem, a person verbally describes their cognitive, perceptual and control behaviours. This theory relies on neurophysiologic and behavioural evidence that speech can take place without any loss in information processing capacity or commensurate performance reduction with regard to signals presented to the visual system. Using the same chest radiography cases, the method of collecting verbal protocols during the interpretation indicated a much higher proportion of search errors and much lower proportion of decision errors than with the Kundel-Nodine gaze dwell time method (Berbaum, Franken et al. 2000). During the verbal protocol procedure, almost half of the observers acted like they created a kind of checklist on the spot to help them generate more systematic verbalization. This may explain the different indications for the source of errors the two methods provided, since "some heuristic method of self-prompting, such as an automatic checklist" might counteract SOS error.

Another experiment (Berbaum, Bradser et al. 2000) demonstrated an SOS effect on test fractures with major, but not minor additional fractures. These experimental results suggest that detection of other fractures is inversely related to the severity of the detected fractures, thus eliminating the faulty scanning as a cause for SOS in musculoskeletal trauma. The average overall search time was significantly reduced by nine seconds when a fracture was added, regardless of the severity of the fracture.

To combat SOS, the design of radiology workstations should evaluate different techniques, to allow for systematic viewing of images, like covering the study in a specific spatial manner (say starting in the upper left hand corner of the image and then

Designing Better User Interfaces for Radiology Workstations 86
proceeding in some set pattern over the remainder of the image) (Rogers 2000). This simplistic solution may prove more like an impediment, since radiologists like to get a 'gestalt' of the entire image/examination, and the radiologist’s eye can be drawn from the very beginning to one or more findings (for reasons not always explicable, but very likely related to the context). A better solution would unobtrusively track the radiologist’s eyes, providing active signalling of the unexplored image areas, and providing notification to prevent the completion of a report without the radiologist seeing all the images.

4.6.3 Understanding the nature of expertise

The nature of expertise can be broadly divided into knowledge and experience. Perception research can look into clarifying situations when errors occur, differentiating between potential causes, such as ambiguous information rather than problems of perception, attention or decision making. Cognitive research, which draws on cognitive psychology, artificial intelligence, philosophy and linguistics, is primarily concerned with characterizing the knowledge of structures and cognitive processes underlying human performance. Using cognition the users of radiology workstations can be ranked as experts, intermediates and novices. User can have different levels of medical competence, corresponding to different levels of training.

Some studies demonstrated specific and predictable differences between novices and experts in terms of perceptual search behaviours:

- experts tend to find lesions earlier in search than novices
- experts have different fixation and dwell patterns
- experts tend to have much more efficient search strategies than novices.

Thus is a radiologist somehow better at searching images than other clinicians or lay persons? Two separate studies compared radiologists and lay persons searching for hidden targets in complicated picture scenes (Nodine and Krupinski 1998). One example of such tasks is finding Nina and Waldo in the ‘Where’s Waldo’ children’s book. These tasks were similar to reading X-rays and searching for lesions because the targets of search were embedded in complicated backgrounds that also had to be searched and interpreted in order to understand the scene, much like tumours in chest radiographs.
Overall, radiologists spent more time searching images for targets, and they also tended to fixate on the target much earlier in search than lay persons did. The radiologists’ scanning patterns suggested a more detailed visual search, which covered less of the image than the layperson's circumferential search pattern. The authors of this study conclude radiology expertise did not positively transfer to the limited art-testing experience. To support their conclusion, Nodine and Krupinski referred to the theories of Osgood, which relates the degree of transfer on the similarity of training and test situations, and Bass and Chiles, which suggest that performance on perceptual tests had little correlations with diagnostic accuracy in detecting pulmonary nodules (Osgood 1956; Bass and Chiles 1990).

A similar experiment, this time involving lesion-detection, was conducted with 16 untrained and 16 trained subjects (Hendee 2002). These results showed large intra-observer variability in both groups, with no significant difference in lesion detection between trained and untrained observers. Since the effect of training proved insignificant in this study, the author suggests that talent may be more important than training, and that detection may be a skill perhaps learned early in life.

4.6.4 Developing aids to image perception

As mentioned in one experiment described in the beginning of Section 4.6, about two thirds of the missed lesions receive prolonged perceptual attention and processing. One could improve these error rates by using perceptually-based feedback, where the eye-position of the radiologist is recorded when searching for lesions. The eye data is then used to circle the areas associated with dwells longer than 1 second. Use of such visual feedback resulted in a 16% increase in observer performance for radiologists looking for pulmonary nodules, compared to just showing the image again without any dwell locations indicated (Nodine and Kundel 1990).

It is possible that ‘computer-aided detection schemes’ improve detection, for the same considerations perceptual feedback did: it focuses perceptual and attentional resources better than the unaided radiologist can do by himself.
4.7  **Key components of the radiologist’s task**

We already studied the radiology workflow through an extensive one-to-one interaction with radiologists from various locations in the United States (Moise 2003). Encouraging radiologist’s feedback from Americas, Europe, and Japan was also received as an early validation of our line of research.

Based on extensive interaction with radiologists from various locations across US, and on the medical imaging literature, we extracted some key features of the radiologist’s task. The primary goal of a radiologist is to produce an accurate diagnostic report in the most efficient manner possible. Identifying various anatomic structures and pathological findings requires mental effort for the visual search, analysis and interpretation of increasingly complex radiological data. When prior radiologic studies are available, the radiologist must also assess the evolution in time of the abnormality. In order to correctly classify a pattern within an image as abnormal, the radiologists often times have to gather and register complementary information from several related images.

We used softcopy reading workflow decomposition and task analysis to come up with the following characteristics of radiologists’ task:

- It is time critical, in that rapid and accurate radiology diagnosis can be vital for saving human lives in hospital departments such as ICU, Emergency, and surgery. Also, radiologists are paid proportionally with the number of studies reported, which provides another reason why the interpretation speed is paramount.

- It involves a visual search using high-resolution displays. A workstation with two monitors is probably the most common solution, but four monitors are also used quite often. Due to the increased physical size of displays, accessing controls from a toolbar located at the top or the bottom of the screen involves significant mouse travel. There is a new trend for PACS vendors to use workstations with heterogeneous displays, such as two greyscale monitors with a spatial resolution of 2500 x 2000 pixels, and a color monitor with 1024 x 768 pixels. For these configurations, workstation controls and text font size should be scales automatically on each monitor, thus adapting to different resolutions.

- It involves a complex workflow consisting of several states. Depending on modality and procedure type, the interpretation task involves multiple stages. Each stage offers a meaningful presentation of information on computer screens.
• It has a highly contingent work flow. The chronological sequence of stages is directly affected by the detection of anatomical abnormalities. Image pairs examined for comparison, and the prior examinations used for inter-study correlation are selected based on the particular question the radiologist must answer, and the real-time detection of lesions. The radiologist can even sometimes change the image acquisition protocol based on preliminary findings, while the study acquisition is still in progress.

• The radiologist has a working memory filled by the task. Often there is one current examination and one or more priors. Each examination can have several series, and each series can have many images. The radiologist has to keep track of which series he or she read already, and where the key images were located inside each series. Some series must be examined several times; such is the case with an axial abdominal CT scan that may require multiple passes corresponding to different window width and level settings. If lesions are detected, their size must be measured and the radiologist must report on the lesion’s evolution relative to the prior examination. The information described above must be stored in the working memory, to be available for quick access. Accessing the workstation controls also requires access to the same cognitive resources, which usually flushes the diagnostic information from the short time memory. Such disruptions and interruption can not only increase the interpretation time, but can also affect the accuracy of interpretation.

To summarize, radiologists’ task is time critical, and involves a visual search on multiple high-resolution displays. A complex sequential work state, highly contingent work flow, and a short time memory filed by task are also specific to radiologists’ work.

4.8 Usability engineering

We used principles from usability engineering to suggest solutions for the placement of controls in order to reduce the disruptions of the interpretation process, and especially interruptions of the visual search.

Finally, we used principles from software engineering to describe an iterative workstation design process based on rapid prototyping and frequent feedback from representative users, as outlined in Section 3.9. Our practical experience with the successful design and deployment of state-of-the-art diagnostic radiology workstations provided invaluable insight for identifying requirements for next generation workstations.
4.8.1 Classification of radiology workstation controls

The radiology workstation tools and features are described in Section 3.4. Based on the task analysis described in Section 3.3 we identified that interacting with the workstation can cause different levels of disruption of a radiologist's diagnostic process:

- visual attention on the features of the control rather than the features of the image under analysis,
- eye movement from the image area of interest to locate a control and guide the hand in executing it, and
- cognitive load to recall, select, and execute controls.

We classified the workstation controls into four classes:

- controls that can be located and operated non-visually and with little cognitive load,
- controls that can be located non-visually but whose operation requires visual monitoring but little cognitive load,
- controls that must be located and operated with visual monitoring but little cognitive load, and
- controls that require significant cognitive load to operate.

4.8.2 Case study: radiology workstation design for the ICU

An exercise to department-specific workflow analysis is presented in Section 3.10, which addresses the main objectives in designing radiology workstations for use in the ICU.

First, we identified the ICU specific conditions in Section 3.10.1:

- In ICU, patients in critical care require intensive monitoring, which often includes non-routine, problem-specific chest radiographs (Shile, Kundel et al. 1996).
- ICU physicians make rapid decisions, based on limited and usually incomplete information, and thus requiring more rapid access to images and radiology interpretations than in most hospital settings.
- PACS can provide faster, more convenient access to radiological images and reports than a typical film-based environment.

Second, we identified the characteristics of workstations used in ICU in Section 3.10.2, such as patient selection using both patient name and bed number, and display modes for 'emergency review', 'historical review', and 'daily overview' (Siegel,
Protopapas et al. 1997). We also noted the frequent situation when ICU physicians perform the first interpretation of radiology examinations (Simon, Khan et al. 1996). In such situations, radiologists should still be consulted on high risk patients or on exams difficult to interpret.

Third, we focused on reliability and speed, as the most important objectives for the radiology workstation in the ICU. Such desirable objectives could be achieved by using workstations with two or four monitors, and a user interface tailored to the main task facing the ICU staff.

We concluded this case study by presenting tools for the automation of most routine tasks, such as ICU oriented HPs capable of automatic loading of relevant priors, automatic image orientation and automatic shutter (Evanoff and McNeill 1997).

4.9 DICOM WG11 and Supplement 60 on Hanging Protocols

Since 2001, through membership of DICOM WG11, the author has been actively involved in the extension of the DICOM standard to provide support for image presentation for different types of viewing circumstances, as described in Section 3.15.4. The soon-to-be-released Supplement 60 to the DICOM Standard proposes solutions for storage, exchange, and query/retrieve of HPs. The objective for DICOM Supplement 60 is to allow radiologists to set up HPs for a variety of reading situations on one workstation, and have the capability to have those HPs available on several other workstations, independent of the workstation manufacturer.

Our major contributions to DICOM WG11 were in the area of image filtering and sorting, and in the consistent reproduction of colour images on different display devices. We investigated colour models used in industry nowadays, both within and outside of the medical community, to determine the most appropriate standards that might be incorporated in the DICOM Grayscale Presentation State to achieve consistent colour presentation independent of display devices. We also worked on the mechanisms used by HPs to group and sort images in order to produce standard presentations. In particular, we worked on formalizing the way image data sets are reordered, to address the needs for image re-grouping and re-ordering for effective softcopy reading.
Based on our research on HPs, and our practical experience with the development of diagnostic workstations, we brought to the attention of the WG11 the following topics which were included in Supplement 60:

- Support for ‘virtual monitors’ and stages. Rather than restricting HPs to communicate only the initial presentation of images, virtual monitors and stages could be used to communicate information about other presentation layouts that are expected to be used during the radiologic interpretation process.

- Prioritization of relaxable image presentation attributes. Rather than using a single binary factor to represent the image layout, the user should be able to specify a preference for preserving the image true size, the spatial resolution, or the layout defined as rows x columns.

- Reconstruction of cross-sectional data. High performance, multi-detector CT scanners can now produce images with 1-2 mm slice thickness, thus generating isotropic volumes of data on x, y, and z coordinates. With increasingly more computational power available on standard, inexpensive computers, it is now possible to interactively choose the parameters for the reconstruction of the two-dimensional images, such as the orientation axis: axial, sagittal or coronal.

- Support for the presentation of flagged images. Flagged images are marked as special and require further review. DICOM does not provide support for flagged images. A referring physician may be interested in a HP that only displays the few images showing the lesion, images that were already flagged and annotated by the reading radiologist.

- Introduction of ‘ranges of values’ to help in better specifying the HP intent. For example, a radiologist interested in the evolution of a patient’s pneumonia could use a HP that automatically prefetches only chest CR images that are less than 6 months old.

- We also compiled a list of additional DICOM group/elements that provide information on HP identification, HP retrieval, and layout, such as image position and image orientation. These are now part of the filtering mechanisms presented in Supplement 60.
5 Proposed solution: HP++

In order to address some HP limitations discussed in Section 3.15.4 we introduce HP++, an extension to current hanging protocols. Our solution focuses on stages and scenario based interpretation to model the interpretation workflow. Other key elements of HP++ include handling heterogeneous data, virtual desktops, and abstract hierarchy organization to allow for approximate query matching. Each of these concepts are described in detail in the following sections.

5.1 Study driven vs. scenario driven interpretation

In a study driven interpretation, the radiologist selects a study from the study list. The selected study is opened, and its images are arranged on the screen according to modality, anatomy, and exam type. More advanced HP implementations could trigger the loading of relevant priors. Only the initial appearance of images on the screen is stored in such a protocol, as defined by series position, layout, W/L, zoom and pan, and annotations. Study driven interpretation is currently used for radiological interpretation, and represents a bottom-to-top approach: starting from this particular study, what is the best diagnosis the radiologist can produce.

To optimise the radiologist’s workflow on softcopy diagnosis, there is a need for scenario based interpretation, which groups data temporally, according to the mental paradigm of the physician. Scenario-based interpretation very nicely handles complex situations, when multimedia information must be pre-fetched using abstract concepts not
explicitly stored in PACS in order to create a sequence of pertinent chronologic events facilitating a temporal visualization paradigm.

In a scenario based interpretation, the selection of data to be displayed for a particular patient depends on the diagnostic protocol, such as 'brain tumour', 'spine interpretation', 'oncology - advanced cancer'. Data from multiple examinations will be automatically selected and displayed. This is an exhaustive top-to-bottom approach especially suitable for complex studies, where screen estate is at premium: starting from the question that needs to be answered, which was extracted from the diagnosis protocol, what is the relevant data the radiologist requires for interpretation, and how to best present it on the current hardware platform?

5.2 **Scenario-based, workflow oriented hanging protocols with stages**

A ‘stage’ reflects the user-workstation interaction required for completing a single step within a complex task, a meaningful view of heterogeneous data. Stages provide context-sensitive navigation, enabling the gathering and filtering of information customized for users within a pre-defined domain. Stages represent our solution to the relatively small size of workstation displays, which provides a meaningful temporal sequencing for image display.

The radiologist would normally advance incrementally from stage N to stage N+1, but jumping to an arbitrary stage should also be possible. A ‘stage’ reflects both the data and the tools required to complete a certain task involved in a reading session. This will provide intelligent filtering of information, alleviating information overload, and reducing the complexity of completing a task, through more manageable and meaningful views. For example, a staged version of the ‘Lumbar Spine’ protocol would include (Valentino, Harreld et al. 1998):

- a first stage that shows the sagittal scout image: requires the ‘Disk labelling’ tool for labelling disk vertebrae;
- a second stage to correlate axial images with the previously labelled scout: requires three dimensional reformatting, W/L adjustment, and image correlation;
- a third stage for reading axial images: requires the ‘Cine’ tool for review of labelled diagnostic axial images, and the ‘Linked series’ tool for comparing labelled diagnostic axial images.

The screen presentation corresponding to each stage can be dynamically modified by the user by changing zoom and pan, W/L, or by adding measurements and annotation. Screen presentation is saved automatically when advancing to a different stage, so when a stage is revisited all the adjustments made by the user to improve the visualization of the current case will still be in effect.

One could store not only the initial display of images, but also the HP stages, i.e. the steps corresponding to this specific diagnostic interpretation process. The reading habits of expert radiologists, who produce best accuracy and interpretation times, can be thus standardized.

Frequently, the radiological examination consists of a routine investigation where no abnormalities are present. Interpreting such studies can be streamlined by using simpler HP, with fewer comparisons and key images required, thus less HP stages. A ‘canned’ or predefined report can be used to act as a macro for such normal findings. The automatic selection of such HPs could be done based on the physician requisition order.

In the situation when abnormalities are present, the sequence of steps (stages) will differ from the normal study. Subsequently, the interpretation flow will adapt according to the anatomical findings detected by the radiologist: investigation of relevant previous examinations may be required, and more comparisons may be needed.

5.2.1 Potential advantages of the use of stages

The main benefits of stages are simpler user-workstation interaction and reduced cognitive workload for softcopy interpretation. These benefits will be especially visible when multiple series must be managed and displayed together for comparison. Once a set of stages corresponding to a certain imaging protocol is defined, the user does not have to keep track of which series were reviewed, or which comparisons should be done next. Staged HPs can be used to emulate the reading patterns of expert users, thus helping to pass on expertise and to potentially improve the performance of less experienced users.
HP++ could become a powerful tool for teaching inexperienced radiologists the best way of reading diagnostic images.

HP++ could also provide better support for radiologists to resume work after task interruptions. Common interruptions occur for injecting a contrast agent to a patient, or for answering questions of referring physicians. The radiologist can remember the context prior to the interruption by reviewing the screen organization from the previous stage, and by listening to the content of dictated partial findings.

If stages are incorporated in the overall design of radiology workstations, interpretation time might be significantly reduced. A good radiology workstation design could initially load the data needed in the first stage, thus minimizing the amount of time the radiologist is waiting for images to be displayed.

### 5.3 Handling heterogeneous data

When PACS is tightly integrated with the radiology information system and the hospital information system, additional pertinent sources of clinical information are accessible to radiologists in digital format, such as radiology reports for previous examinations, patient's complaints, admitting diagnoses description, laboratory test results, requested procedure description, study description and reason for the requested procedure. For example, a pulmonary timeline scenario will display the pulmonary function, the chest X-ray, and clinical lab results, such as arterial blood gases.

HP should include a first stage for the presentation of auxiliary information, such as previous reports, diagrams, and electrocardiograms to the radiologist. This stage could help create a better context for the reading session, complementing information currently displayed on screen, such as the patient's age, gender, and number of prior examinations.

We believe radiologists can become effective clinical consultants if access to pertinent sources of clinical information is provided at the time of dictation. With integrated clinical information access, radiologists' interpretation not only comments on morphologic findings but also can enable evaluation of study findings in the context of
pertinent clinical presentation and history. HP++ links heterogeneous data in a way that permits users to look at medical information in a single unified view.

In February 2003 the author asked by email the opinion of three prominent radiologists and specialists in softcopy reading: Dr. S. Horii and Dr. H. Kundel from University of Pennsylvania, and Dr. N. Strickland from Hammersmith Hospital in London, United Kingdom (see her quote in Section 5.6). The question I asked was: “Could you please comment on the trend of the radiologists becoming an effective clinical consultant with access, at the time of dictation, to pertinent sources of clinical information, such as prior radiology reports, pathology data, laboratory data, history and physicals, clinic notes, and operative reports. Is this extra information and the associated increase in the time to dictate a case worth it?” The response from Dr. Steven Horii, a radiologist from University of Pennsylvania, was that:

Many radiologists already consider themselves to be consultants. Most radiology request forms are actually titled, “request for radiologic consultation.” Certainly, if we find something on the images, we try to give a concise diagnosis as possible and we will make use of other information available. Most well-trained radiologists will do this as a matter of course. An example phrase in a report might be: “In view of the patient’s history, the findings favor an infectious over neo-plastic etiology.” We also frequently comment on progress of disease; whether improving, stable, or getting worse (Horii 2003b).

Another radiologist, Dr. H. Kundel responded:

I think that the more accurate information the radiologist has, the less the likelihood of an error and the greater the likelihood of a correct diagnosis... I have been an advocate of integrated medical information systems for many years. I would have liked to have access to a patient’s past medical history and old images whenever I read cases. From the frame of reference of the radiology department this is not cost effective, productivity is the name of the game... From the frame of reference of the patient who acts as if there is an infinite supply of money, it is effective (Kundel 2003).

5.4 Virtual desktops

Due to limitations of the screen real estate, radiologists often change the layout and the content of the screen(s) sequentially during the reviewing process. In order to
model the entire screen presentation during radiological interpretation, a whole sequence of HPs ('virtual desktops') could be used for sequential review. Virtual desktops represent 'logical' viewports (Ghosh, Andriole et al. 1998), which is independent of the actual number and properties (such as resolution, support for colour, refresh rate) of screens available at the system. The user can control which desktop is currently active or displayed.

For instance, in digital mammography radiologists examine right and left cranial caudals (CCs) at full resolution, then right and left mediolateral obliques (MLOs) at full resolution. These steps are necessary because unlike film the electronic display resolution is not capable of displaying all four images at full resolution at once. If the HP communicates only the initial presentation of right and left CCs, then the HP does not capture what the standard softcopy or hardcopy (film) presentations would display (CCs and MLOs). However, by defining virtual desktops, the intent to present additional material (in this case the MLOs) can be indicated.

Virtual desktops could be extended to work with focus plus context screens (Baudisch, Good et al. 2002). Focus plus context screens are wall-size low-resolution displays with an embedded high-resolution display region. They provide a single, non-distorted view, which is essential for the display of medical images. For example, when tile mode is used to display cross-sectional studies, the cluster of images currently reviewed by the radiologist could be displayed on the smaller, higher resolution screen, while the context is shown on the wall-size, lower resolution display.

5.5 Type abstraction hierarchy and approximate matching

In a previous paper we presented a HP selection paradigm based on the following hierarchy: modality, body part, and procedure type (Moise and Atkins 2002c). The search for a matching HP will start on the deepest level of the tree. If no match is found the

---

9 Viewing screening mammography studies on film requires three primary comparisons: the same projection, such as the new right CC compared with the new left CC; the new and old views of the same breast and same projection, such as the new right MLO and old right MLO; and the two projections of the same breast for the same study, such as the new right CC and new right MLO.
search would go up in the tree, eventually reaching the modality level. While this type of HP structure is currently found on most advanced radiology workstations, there are significant limitations associated with this fixed priority scheme:

1. It is not always suitable to use this scheme for querying; for example, the solution proposed by DICOM WG11 allows for some degree of flexibility in defining the intent for the hanging protocol, while providing a precise structure for query matching. We take this concept to the next step, by allowing for approximate query matching. For example, aliases could be used to compensate for inconsistencies in series labelling. With abstract categorization, multiple potential sources of values can be used. For example, for the abstract category PROCEDURE_INTENT one can use values from physician's current procedural terminology codes\(^\text{10}\), The Systematized Nomenclature of Medicine (SNOMED) codes\(^\text{11}\), International Classification of Disease, 10th revision (ICD-10) Procedure Codes\(^\text{12}\), or Local Codes.

2. relies on values are associated with specific DICOM Composite Information Object Definition (IOD) attributes, thus reducing the potential for HP inter-changeability. For example, ‘Anatomic Structure’ (0008, 2208) - obsolete: DICOM recommends use of SNOMED, ‘Body Part Examined’ (0018, 0015) is optional, and the ‘Anatomic Region Sequence’ (0008, 2218) is only supported by some IODs.

3. poor filtering and sorting: DICOM Study and Series Unique Identifiers (UIDs) are assigned by imaging modalities, and so depend on details of assignment policies of these devices. In addition, the order of images is not fixed, especially since many modalities provide each image as a separate instance rather than providing ordered multi-frame images. However, radiologists usually require images to be grouped and sorted in standard presentations. Thus, ‘raw’ DICOM data, as received by workstation or archive, is often not suitable for radiologists’ purposes. It is therefore useful to process selected images by filtering them to create subsets (‘image set sequences’) that can be displayed in different groupings than those indicated by DICOM study and series UIDs.

4. no support for constraints: in many situations, the radiologist must compare one or more images. It is important to be able to specify that the corresponding data sets these images belong to must be displayed simultaneously and at adjacent screen locations, in order to facilitate image comparison.

\(^{10}\) PHYSICIANS’ CURRENT PROCEDURAL TERMINOLOGY (CPT). A systematic listing and coding of procedures and services performed by physicians and other clinicians that is widely used for coding in billing and payment. The CPT listing, developed over decades by the American Medical Association, assigns a unique code to each medical procedure, with modifiers to refine the descriptions of the services provided (http://www.aacap.org/clinical/cptgloss.htm).

\(^{11}\) The Systematized Nomenclature of Medicine http://www.snomed.org/

\(^{12}\) International Classification of Disease, 10th revision http://www.who.int/whosis/icd10/index.html
5. support for classes of equivalence, required to adapt existing, viable HPs various hardware configurations and user preferences.
6. no support for scenarios, stages or ‘virtual monitors’
7. no support for heterogeneous data

In order to address 1-5, our proposed solution takes advantage of type abstraction hierarchy introduced by Chu in 1994 (Chu and Chen 1994; Chu, Chiang et al. 1996). In the next paragraphs we will apply several concepts from of type abstraction hierarchy, such as attribute relaxation order and non-relaxable attributes, to demonstrate how HP++ can be used to address a very common limitation of current HPs: dependency on the hardware configuration used when during HP definition.

Let us consider the example of a HP for ‘Generic CT 1024x1024’, defined by radiologist X on a radiology workstation with one display (1024x1024 pixels). This HP describes a 2x2 layout. Let us assume that radiologist X is now reading a CT examination on a workstation using one 2560x2048 monitor. There is no HP defined by radiologist X for this hardware configuration, so we will attempt to adapt the ‘Generic CT for single 1024x1024 monitor’ HP to the new configuration. In order for the HP adaptation to occur, one or more attributes of ‘Generic CT for single 1024x1024 monitor’ must be relaxed. For simplicity, let us enumerate the following candidate attributes:

- **layout**: preserving a 2 x 3 layout on a monitor with a much higher resolution will generate images too big, with no benefit for the radiologist;
- **image resolution**: since each image is 512 x 512 pixels, 20 images could be displayed at full resolution on a 5 x 4 layout
- **image ‘true’ size**: matching the size of anatomical structures to its real size dimensions; requires calibration information for both the images and the display, and it is mostly used for the display of X-ray of bones.
- **image apparent size**: the dimension in cm of the image on screen.

The ‘instinctive’ approach to this particular HP adaptation is to relax the layout to a 5 x 4 format, thus continuing to display images at 100% resolution. However, it was reported (Beard, Smith et al. 1996) that images displayed in a 5 x 4 layout seemed too small for the radiologists, so a 4 x 3 layout should be used instead. To allow for such solutions, we propose embedding the attribute relaxation order into the HP structure, to preserve the user’s preference, such as ‘preserve image apparent size’. The role of the
image size on the observer's ability to detect lesions, such as lung nodules on a CT scan, is well demonstrated (Seltzer, Judy et al. 1998). So we also suggest defining non-relaxable attributes, such as 'true size' for orthopaedist, or minimum monitor luminance and resolution for mammography interpretation.

5.6 Potential limitations for the use of HP++

The clinical unpredictability of many imaging studies could reduce the applicability of predefined HPs. While a rigid scanning protocol is generally used, often times the image acquisition procedure is changed while the examination is still in progress, according to anatomical findings within the study. When asked to comment on this subject, Dr. N. Strickland wrote in an email from July 4, 2002:

"New MR sequences are being developed all the time and it is very common for us to modify the series acquired according to the clinical condition, and to modify the sequences acquired according to findings within the study in real time while the examination is in progress (Strickland 2002).

Some inconsistencies, such as misspelled series/protocol names, series acquired in a different order, and incorrect orientation of the images, can be compensated for automatically. However, using staged HP assumes a certain level of consistency for the acquisition protocols that may not be always possible. For example, poor user interface on imaging modalities produces images with incorrect parameters (Guld, Kohnen et al. 2002). Also, certain DICOM Composite IOD attributes (such as 'Anatomic Structure', 'Body Part Examined', and 'Anatomic Region Sequence') used for HP querying and selection are obsolete, optional, or only supported by some IODs.

Current radiology workstations use study-level DICOM information for hanging protocols selection. Information stored at series level, such as number of series and images, type-pulse, and series description, and image level, such as orientation, slice thickness/distance, use of contrast, phase, and exposure, is not available to help refine the HP selection. The solution to this problem may come from parsing the DICOM headers when the study is received from the modality, and storing this 'metadata' for future reference. Access to this metadata could facilitate HP query sequence matching, using variable combinations of abstract concepts, such as modality, anatomy, procedure code,
and study intent. The metadata could also be used for patient record augmentation and conceptual patient clustering. Support for approximate query matching can be implemented by providing attribute relaxation order (e.g. image size vs. layout), and non-relaxable attributes (e.g. true size for orthopaedist).
6 Pilot experiment

6.1 A radiology look-alike task

6.1.1 Motivation

Medical image interpretation is inherently uncertain in nature, partly because two dimensional images drawn from the real three dimensional world are an incomplete representation.

Identifying various anatomic structures and pathological findings requires mental effort. The task of interpreting medical images is complicated, involving simultaneous processes of searching, perception and decision making. Radiologists have stringent requirements of accuracy, confidence, and speed. A crucial requirement for increasing radiologist productivity is to reduce disruptions of their analysis, particularly disruptions of visual search. Accessing the controls of current radiology workstations produces considerable disruption of the visual search, which may lead to reduction in the amount and type of information processed, the methods by which human incorporate, use, and/or ignore computer prompts. To prevent this disruption, frequent tasks such as navigation and image manipulation should require little or no visual and cognitive resources. Dr. Harold Kundel, a radiologist at Hospital of the University of Pennsylvania wrote:

Most of the people that I know would like an optimal image to appear that does not require much manipulation and whatever manipulation is required is very simply done. This is the problem of walking and chewing gum at the same time. If you are paying attention to manipulating the
image you not paying attention to the reading task. So somehow the manipulation, rove, zoom, and so on has to become a background rather than a foreground activity (Kundel 2003).

6.1.2 Real-life radiology scenario

In order to implement HP++ instances to be used in real-life scenarios, radiological domain knowledge is essential. Acquiring such expertise would involve several weeks of work with radiologists, just for the development of a single hanging protocol family associated with a specific radiological interpretation task, such as 'MR Brain Tumour', or 'CT Spine'. The next step would require the iterative development of an application to emulate softcopy reading using HP++. Unfortunately, such endeavour would be too expensive for us: there is no medical facility with a big radiological department accessible for our research, and the development of a limited functionality radiological workstation would take too long to be practical for our research. Consequently, we decided to support our research hypothesis with results from different, but related domains. Transferring our results to softcopy interpretation will be possible by capturing the characteristics of a typical radiological interpretation task.

A representative radiological task involves the radiologist reading a current chest computed radiography (CR) study with a post-anterior (PA) and a lateral (LAT) projection, when a prior study (also containing a PA and a LAT image) is available. Figure 32 below shows a 1999 chest exam of a patient, while Figure 33 shows the prior exam of the same patient.

Designing Better User Interfaces for Radiology Workstations
The radiologists’ task consists of detection of abnormalities present in the new study, in the context of additional information available from the prior study. If common abnormalities, such as chest nodules or broken bones, are detected, radiologists must characterize their evolution in terms such as “lesion became smaller,” or “changed from benign to malign.” Radiologists must also mention if an abnormality from the prior study is not present in the current study, or if any new abnormalities are visible in the new study that were not present in the prior study. Radiologists can usually access the report associated with the prior examination, which simplifies their understanding of the prior study.

Designing Better User Interfaces for Radiology Workstations
However, often there are situations when:

- no radiologic report is available, which occurs when the report is lost or not-accessible, or
- the abnormality is not mentioned in the radiologic report for the prior study,
- the report includes a false negative, or
- abnormality was too difficult to detect at that time, or
- abnormality was masked by other anatomical structures

In any of the situations described above, radiologists must interpret both the old and the current study, with no a priori indication about the presence and location of any abnormalities.

In many clinical environments, radiology workstations used for chest CR interpretation use two high-resolution greyscale monitors. Consequently, two CR chest images can be displayed at full resolution at the same time. In order to read a chest CR study, with two images (PA and LAT), when an old examination with another two images is also available for comparison, the radiologist has to interact with the system in order to select which 2 images are to be displayed on the screens. This image navigation is usually done using workstation controls in the form of thumbnail images, which represent scaled-down copies of the original images. The use of thumbnails allows the user to get visual feedback on the kind of image corresponding to a particular control. However, the resolution of thumbnails is too low to allow the differentiation between two images of the same type, such as the differentiation between an old PA image and a new PA image. These thumbnails are labelled with the study acquisition date and time.

We considered four possible relevant comparisons a radiologist may perform while reading a CR chest with a prior study. Using the principle of HP++, we defined the four stages for the interpretation scenario described above, which are summarized in Table 13.
Table 13: Four stages for interpretation of a CR chest exam, when a prior study is available for comparison

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage description</th>
<th>Full-size image displayed on left monitor</th>
<th>Full-size image displayed on right monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Current study</td>
<td>PA_current</td>
<td>LAT_current</td>
</tr>
<tr>
<td>2</td>
<td>Prior study</td>
<td>PA_prior</td>
<td>LAT_prior</td>
</tr>
<tr>
<td>3</td>
<td>Comparison PA images</td>
<td>PA_current</td>
<td>PA_prior</td>
</tr>
<tr>
<td>4</td>
<td>Comparison LAT images</td>
<td>LAT_current</td>
<td>LAT_prior</td>
</tr>
</tbody>
</table>

We chose to model this representative case study from radiographic interpretation because:

- Chest examination is a very common form of examination, especially after the implementation of screening programs that periodically check on people with a high-risk, but no symptoms. Computed radiography accounts for about 70% of the radiological exams clinically performed (Reiner 2002).
- We used CR as it is simple, with only four images, whereas the interpretation of CT or MR studies is more complex and involves handling several sequences, each composed by many images.
- These four HP++ stages map naturally on the radiologist’s interpretation workflow.

### 6.2 Purpose

A pilot study was conducted prior to the main user study, with approval from the Simon Fraser University Office of Research Ethics (see Appendix H: Approvals from the Office of Research Ethics). The goal of the pilot experiment was to analyse the performance of subjects engaged in a radiology look-alike conjunctive visual search (see Section 6.4 on Conjunctive visual search) under two different interaction techniques. We will refer to the first interaction technique as Free User Interface (FUI). This interaction technique is found in many current radiology workstations. The second, new, scenario-oriented interaction technique will be referred to as Stages, since it is based on the concept of staging.
We videotaped, collected comparative information and also performed eye gaze tracking while subjects were performing the task, to determine where subjects are gazing during the task. Under both interaction techniques conditions, the computer screen had similar screen layouts consisting of a left viewport, a right viewport, the background, and the navigation controls which differed under the two conditions.

A second goal for the pilot was to tune the experimental conditions for the main user study. The pilot provided feedback on the complexity of the visual search task, the time required for each trial, and the variables affecting the accuracy of the interpretations.

6.3 Hypothesis

Our hypothesis is that using workflow oriented stages streamlines the radiologic interpretation task, which leads to:

1. shorter completion times;
2. less user interaction;
3. improved productivity (more studies performed, and fewer errors)
4. fewer disruptions of the visual search.

Using stages not only has the potential of saving the time the user spends interacting with the workstation, but it can also alleviate the radiologist’s workload and cognitive overhead caused by:

- absence of a skilled technician to handle the initial presentation of images prior to review by radiologist,
- workstation manipulation tasks and workstation constraints not found on film.

When compared with Stages, FUI not only adds a few extra clicks to the interaction, but also reduces productivity by causing disruptions of the scenario analysis, particularly disruptions of visual search. We hypothesised that the improvement in time achieved with Stages will be more than just the time needed to perform the extra clicks with FUI.
6.4 Conjunctive visual search

Visual search tasks are those tasks where one looks for something. For example, a radiologist looks for pulmonary nodules on a chest radiograph. In a standard visual search, subjects look for a target item among some number of distractor items. The total number of items in the display is known as the set size.

Some searches allow for the processing at once ('in parallel') of all the items from a scene. The target item seems to 'pop out' of the display, and the quick response time is independent of the set size. Consider a search for a red item among green distractors. As has been shown many times, the number of green items makes very little difference (Nagy and Sanchez 1990). Either red is present or it is not. The usual inference in the visual search literature is that these results reflect an underlying parallel search. Apparently, all items can be processed at once to a level sufficient to distinguish targets from non-targets. The red item, if present, 'pops out' and makes its presence known.

In other searches it seems that the participant is forced to study each item individually until the target item is found. In these cases the target item does not pop out, and search time increases with the number of distractor items. Searches yielding this pattern of results are usually called serial searches because the pattern of results is consistent with a random, serial self-terminating search through the items.

Anne Treisman proposed a division between parallel and serial visual searches in her original Feature Integration Theory (Treisman and Gelade 1980). Triesman's proposal was that many feature searches were parallel searches and that everything else required serial search. Feature searches are searches where the target is distinguished from distractors by a single basic feature like colour, size, or motion. 'Everything else' included searches for targets defined by conjunctions of features. For example, in a search for a big red square among small red and big green squares, the target is defined by a conjunction of colour and size. Neither the size nor the colour feature alone defines the target.

Most searches in the real world are not searches for stimuli defined by single basic features. They are searches for stimuli that are defined by conjunctions of two or
more features. One does not look for ‘banana’. One looks for an apple that is some conjunction of yellow, elongated, and with the size of a banana. We designed our targets as conjunctions of several features, as described below.

6.4.1 Target description

The subjects were required to find a target on a grey background. Initially we considered presenting to our subjects two different 3D projections of a familiar 3D object with an embedded target, such as finding seeds in two projections of a water melon. This would have been similar to a radiologist correlating the information from the PA and LAT images, the two orthogonal projections in the case of a chest X-ray. According to Dr. H. Kundel, registering information from both projections is specially useful for difficult diagnoses (Kundel 2003):

Fortunately for radiologists most diagnoses are easy. The radiologist and the lateral projections contribute when the diagnosis is hard. Very hard diagnoses are rare but they add value and misses can be costly.

However, radiologists are trained to integrate two dimensional projections into a mental 3D anatomical model of the patient. We were concerned about the great variation in our subjects' ability to understand a 3D structure from orthogonal projections, even when simple objects, such a watermelon, was imaged. Basically, we wanted to avoid interference on our dependent measures due to the perceptual load with 3D stimuli.

Controlling the complexity of the 3D stimuli would have been a difficult task, and our dependent measures were subjected to interference from an observer’s ability to integrate 2D projections of a 3D volume. Research from computer aided design showed that learning to use orthographic projections is challenging and requires practice and help: it seems that at first people need a 3D view to help them orient the 2D views, but once they received enough practice to become comfortable using projection, the 3D view could become less important (Pillay 1984; Osborn 1992).

Woods talks about relating multiple views in a more general sense, and not just 3D/2D specifically (Woods 1984). When people have to make sense of projections of objects from different angles, they tend to use 2 strategies, either separately or combined.
(Osborn and Agogino 1992). Tory also suggested these 2 strategies (Tory 2003), but only qualitatively as it was difficult to separate the two strategies and force people to use one or the other:

1. Mental rotation: mentally rotate the views until they align
2. Feature matching: match up unique features in the views to figure out the relationship.

For the design of our stimuli we consulted a psychologist from the University of British Columbia, with expertise in vision and visual attention. Our choice of stimuli was designed to remove the first strategy and to force people to use the second strategy. We opted to remove a confounding factor by focusing the experiment on the second strategy only. However, this solution may partially reduced our ability to generalize our results to a radiology task where the mental rotation strategy likely comes into play.

To summarize, we wanted to avoid confounding factors relative to the selection of the 3D shapes and their projections. We wanted the detection of a target to put a significant demand on a subject’s attention and working memory, thus stimuli requiring a simple, one-feature matching strategy from our subjects had to be avoided. We preferred to force our subjects to build at least a rudimentary understanding of the complex nature of the scene.

Consequently, we decided to use abstract targets and images rather than radiology images, since the subjects we planned to use were computing science students with no background in radiology. In our experiment, the target is an item with two circles, of the same size, half-split along the same vertical or horizontal diameter, and half-shaded. Three examples of targets are presented in Figure 34. A search using a conjunction of features was required for the discrimination of targets from distractors, since the target was a conjunction of features in the distractors.
Identifying the target on a single image was too easy. So we increased the complexity of the trial by presenting the targets in such a way a subject was able to discriminate a target from a distractor solely by integrating the information from two related images. To achieve this, the target was incompletely revealed to the user due to partial occlusion. A similar occlusion occurs in radiology frequently due to anatomical structures shown as bright areas in the image, which overlay the lesion. Such is the case of a barely visible lung nodule hidden behind a rib on a chest CR, or a liver tumour hidden behind a blood vessel.

The occlusion was simulated in our stimuli with the introduction of a ‘wild-card’, which forced our subjects to register information between the two images of a study. A wild-card was used to represent the disc divider, an important characteristic feature of a target. The disc divider was occluded by a disc with a uniform fill, which could hide a disc divided either vertically or horizontally. The user must find on a related image the actual instantiation of a wildcard.

Note that for instantiation, the top disc on the first image is compared with the top disc on the second image, and then the process is repeated for the two bottom discs. Only the orientation of the divider is important. It does not matter which half of the disc (top or bottom for a horizontal divider, and respectively left or right for a vertical divider) is grayed-out.

Depending on the orientation of the occluded disc divider, a wild-card could either instantiate into a target, or into a distractor, as shown in Figure 35 and Figure 36,
respectively. A third situation, also corresponding to a distractor, occurs when the wild-card does not instantiate into a divided disc, as illustrated in Figure 37.

![Figure 35: The same target is incompletely presented on two different images. Registration is required for solving the 'wild card' into a target. The wild card instantiates in a disc with the proper divider orientation.]

![Figure 36: The wild card can instantiate in a disc with incorrect divider orientation, thus representing a distractor.]

![Figure 37: Distractor, because the two wild cards do not transform into a split disc.]

It is important to note a potential target always had a wild-card. Consequently, for every potential target containing a wild card a subject had to register complementary information from the two images of the same study. Our subjects were instructed to never try to find a target by registering information from two images of different studies.
6.5 Stimuli

To simulate the radiologist’s follow-up on a radiographic examination, we introduced a time dimension by presenting to our subjects two instances of the same scene, corresponding to different time moments. In conclusion, for the experiment we asked our subjects to detect the target, and then track its evolution in time.

An example of stimuli used is presented below. Figure 38 and Figure 39 show the first and respectively second image from the first study. The target is free floating in the bottom left of each image. Figure 40 and Figure 41 show the two images from the second study. The target is no longer present in the second study. More examples of our stimuli are shown in Appendix C: Examples on how various factors affect the trial complexity.

![Figures 38, 39, 40, 41](image)

6.6 Procedure

The pilot involved 2 male and 2 female subjects, who performed two blocks of trials, one block for each interaction technique. The two blocks contained the same 10 trials, presented in the same order (1 to 10) to our subjects. We counterbalanced for the
order in which the interaction techniques were presented to our subjects, meaning that one male and one female subject started with Stages and then performed using FUI, while the other male and the other female subject used first FUI, and then Stages.

Once the subjects were comfortably sitting, instructions about the task were given using several training steps presented on the computer screen. Each training step was followed by a short practice session, where the subjects’ understanding of the recently learned concepts was tested. For more information on our training procedure, refer to Appendix D: Training script for the user study.

The three training steps are described below:

- In the first training step, the concept of a ‘target’ was introduced to our subjects. A short practice session asked the subjects to identify the targets from a set consisting of targets and distractors.

- The second training step asked the subjects to look for a target in a ‘study’ containing two related images. This time one of the discs composing a target was replaced by a ‘wild card’. The subjects had to find on the second image what the wild card stands for, in order to differentiate a target from a distractor. A second practice session asked the subjects to identify if there is a target in the each of the three individual studies used for training.

- The third training step asked to look for a target in two related studies (a ‘new’ study, and an ‘old’ study). Each individual study had at most one target. If a target was found in both studies, it must be the same target (except that the size of the target may remain the same, or increase, or shrink). In the practice session number three, the subjects had to:
  - report if there was a target in the new study;
  - report if there was a target in the old study;
  - in case a target was found on both studies (so it was the same target), report how the target’s size evolved (remained the same, or grew, or shrunk).

After learning about their task, the subjects were introduced to the application used during the experiment. The layout of the screen consisted of a left viewport, a right viewport, the controls used for image selection, and the controls used to start/stop the current trial, as illustrated in Figure 42.

The application allowed for the simultaneous display of only two images. The subjects had to interact with the system to see:
the two images from the first study, and then
the two images form the second study, and in case a target was present in both studies
a comparison between one image from the first study, and one image from the second study.
The selection of the two images to be displayed on screen was done using four thumbnail-size controls located in the top-left side of the screen.

The user interface for Stages, our new, scenario-oriented interaction technique, is shown in Figure 42. For Stages, each of the four controls corresponded to a predefined pair of images as shown in Table 14. A single click on one of the controls resulted in changing the images from both viewports at the same time. For example, by clicking on the left-most thumbnail, the two images corresponding to the first study will be displayed on screen: the PA on the left viewport, and the LAT on the right viewport.

Table 14: In Stages, each control corresponded to a pair of images to be displayed on screen.

<table>
<thead>
<tr>
<th>Control number</th>
<th>Left Viewport</th>
<th>Right Viewport</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PA&lt;sub&gt;1&lt;/sub&gt;</td>
<td>LAT&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>2</td>
<td>PA&lt;sub&gt;2&lt;/sub&gt;</td>
<td>LAT&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>3</td>
<td>PA&lt;sub&gt;1&lt;/sub&gt;</td>
<td>PA&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>4</td>
<td>LAT&lt;sub&gt;1&lt;/sub&gt;</td>
<td>LAT&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
</tbody>
</table>
In FUI, the screen has a two-up fixed layout. One toolbar containing the image thumbnails was used for the independent selection of the image displayed in each one of the two screen locations, as shown in Figure 43. Since four distinct images can be displayed at each of the two screen locations, the user can create a total of sixteen screen combinations. For FUI, the four controls corresponded to the four images to be searched for targets, as shown in Table 15. A two-step interaction was required to change the image in each viewport: first the user had to select the viewport (either left or right), and then the control corresponding to the image they wanted displayed in that viewport. Consequently, to change both images on screen, four clicks were required\(^\text{11}\).

\^\text{11}\ The application remembered the last viewport selected on the screen. One could save the click used to re-select the last viewport. Since there always was a selected viewport, one could change both images on the screen in only three clicks.
Figure 43: The two images from study two are shown from a trial with FUI. The target is in the cross.

Table 15: In FUI, each control corresponded to a single image.

<table>
<thead>
<tr>
<th>Control number</th>
<th>Associated image</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PA₁</td>
</tr>
<tr>
<td>2</td>
<td>LAT₁</td>
</tr>
<tr>
<td>3</td>
<td>PA₂</td>
</tr>
<tr>
<td>4</td>
<td>LAT₂</td>
</tr>
</tbody>
</table>
For the completion of each trial, our subjects had to perform the following steps:

Before each trial, an empty screen was displayed. The only button enabled was ‘Start Trial’. A screenshot corresponding to this step using Stages is shown in Figure 44.

![Figure 44: An empty screen shows no stimuli. The only enabled button is ‘Start Trial’.

When ‘Start Trial’ was clicked, a timer started to record the response time. ‘Start
Trial’ button became disabled. The ‘Stop Trial’ button and all the thumbnail buttons
became enabled. The two images corresponding to the first study are automatically
displayed in the two viewports as shown in Figure 42. Subject had to find if a target was
present in this first study. If a target was found, the subject had to say “First study: here is
the target” and point at the target with the mouse cursor. If no target was found, the
subject had to say “First study: no target.”

After the first study was interpreted, the subject had to examine the second study.
The two images corresponding to the second study had to be displayed in the two
viewports. This was achieved by simply clicking on the second thumbnail, if Stages was used. With FUI, four clicks were required to produce the same result, as shown in Figure 43: first the subject has to click on the left viewport, and then on the third thumbnail to bring down the PA image from study two. Then, the subject must click on the right viewport, and then on the fourth thumbnail to bring the LAT image from study two. The subject must search for a target in study 2. If a target was found, the subject had to say “Second study: here is the target” and point at the target with the mouse cursor. If no target was found, the subject should have said “Second study: no target.”

In the next step, an image from each study was displayed on screen in order to assess the evolution of the target’s size. It corresponds to Stage 3 from Table 13 (shown in Figure 45, or Stage 4 from Table 13. This step is optional, and only makes sense if a target is found in both studies. However, if the subject was confident in assessing the target’s evolution, such is the case with many trials where the target size remains the same, this step is skipped by the subject in order to reduce the response time.

Figure 45: This layout corresponds to Stage 3 of a trial, showing the PA images from study 1 and 2.
The subject had to click on 'Stop Trial' after finishing the trial interpretation. The timer associated with the trial was stopped automatically, and the empty screen was shown again. The 'Stop Trial' button and all the thumbnail controls became disabled, and the 'Start trial' button became enabled. The subject could not go back to see the images from the trial that was just completed. The system was back to step 1, ready to start another trial.

The steps described above are always required in order to properly and optimally complete a trial. However, additional steps can be introduced by a user, which derail from this optimal interpretation strategy. For example, a subject may choose to revisit one of the previously completed steps. It is also possible for a subject to accidentally stop a trial immediately after interpreting the first study. Such an event will generate an incorrect interpretation, as the second study was completely missed.

Our subjects were verbally instructed to be as accurate as possible in finding the target, and reporting the target change in size whenever applicable. We also told our subjects that we will record the completion time of each trial, and encouraged them to work as fast as possible, while not affecting the accuracy of their interpretation.

### 6.7 Independent variables

The only independent variable was the interaction technique, with two conditions: Stages and FUI.

### 6.8 Dependent measures

For the pilot study, we logged the user interaction with the workstation, which gave us information about the Response Time and Number of Mouse Clicks per trial. Using video analysis, we determined the Accuracy of Interpretation for each individual trial. Using an ASL Model 504 remote eye-tracking system we collected Eye Gaze data from our four subjects, as an indication of the visual search pattern. We also measured our subjects' cognitive overload using the NASA Task Load Index (TLX). For a full description of the TLX and the scoring procedure, refer to Appendix B: NASA Task Load Index (TLX).
6.8.1 ASL 504 Eye Tracker

Applied Science Laboratories (ASL) Model 504 presented in Figure 46 is an eye tracking system used where the stimulus presented to the subject is restricted to a planar surface such as a computer monitor and where head mounted optics are not desirable. The system allows the subject approximately one square foot of head movement which eliminates the need for head restraint. The system uses bright pupil technology, giving superior capture and contrast. The system emits near-infrared light invisible to the human eye, to create bright pupil and cornea reflection (CR) signals. Since the system measures the distance between pupil centre and CR, it is sensitive to rotation, but not translation.

Figure 46: ASL 504 remote eye-tracker

For data output, the Model 504 is designed to measure a subject's eye line of gaze with respect to a single stationary surface in the environment, such as the computer screen. Recorded data include time, x and y eye position coordinates and pupil diameter, with a frequency of 60Hz, an accuracy of 0.5 degree visual angle, and a resolution of 0.25 degree visual angle. Eye position coordinates correlate to specific areas on the screen being viewed.

Before an application can obtain points of regard from the eye tracking system, a calibration procedure for each subject must be performed. This procedure establishes a mapping between the raw measurements that the eye tracker collects and the screen coordinate system. The procedure involves recording the displacements between the pupil...
and the CR responses of a subject while sequentially fixating on each point on the calibration grid, as shown in Figure 47.

6.9 Pilot experiment results

6.9.1 Response Time (RT)

In Figure 48 we present the total time for each block of trials, both in the Stages and in the FUI condition. For 3 of our subjects, the RT was smaller when Stages was used, compared to when FUI was used. After reviewing the video tapes of the experiment, we think the subject was faster for FUI because of the learning effect, as he started the task with Stages.

The average RT for a block of 10 trials was 371 s for Stages, and 445 s for FUI, which means that Stages required an average of 17% less time than FUI.
6.9.2 Mouse clicks (MC)

The number of mouse clicks for each subject is shown in Figure 49. As expected, the interaction technique had a significant effect on the number of clicks required to complete a block of 10 trials, with MCavg = 28 for Stages, and MCavg = 79 for FUI.

6.9.3 Interpretation errors

Our subjects were instructed to be as accurate as possible in their diagnosis, as their primary task involved the detection of targets. Completing each trial in the shortest possible time interval was a secondary requirement. This is the reason our hypothesis
made no references to the distribution of errors between the two interaction techniques: we traded time for accuracy. However, 7 trials contained interpretation errors when FUI was used, and 5 trials had interpretation errors when Stages were used. The same two subjects generated all the errors for both Stages and FUI.

6.9.4 Eye gaze

We wanted to assess the impact our stage-oriented ergonomic layout of controls had on subjects' eye-gaze distribution. Under the FUI condition, the total eye-gaze time spent over the controls represented on average 17% of the response time. Considering our controls are a simplified version of radiology workstation navigation controls, our results are consistent with previous eye-tracking research involving radiological interpretation of bone-trauma computed radiographs which reported a 20% distribution of eye-gaze over the workstation controls (Krupinski and Lund 1997). Under the Stages condition, only 6% of the total eye-gaze time was spent on the workstation controls, as illustrated in Figure 50.

![Average % eye-gaze distribution](image)

**Figure 50: Average distribution of eye-gaze**

In Figure 51 we present the distribution of total eye-gaze time over the main screen elements of our interface, while Figure 52 shows the distribution of eye gaze for each subject, under each interaction technique condition. It is interesting to note the equal amount of time (72s) spent on average on the right viewport under both IT conditions. For the left viewport, our subjects spent on average 82s with Stages, and 92s with FUI.
6.9.5 NASA Taskload Index (TLX)

Workload is a multi-dimensional psychological construct measuring the subjective experience of work that results from the mental actions performed while perceiving and processing information and executing a response. The TLX measures workload on 6 different dimensions: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration. TLX is used to create a picture of the amount and type of mental workload a user experiences during task performance (Hart 1987). For a full description of the TLX and the scoring procedure, refer to Appendix B: NASA Taskload Index (TLX).

TLX is a two-part evaluation procedure consisting of both weights and ratings. The first requirement is for each subject to evaluate the contribution of each factor, or the factor’s weight to the workload of a specific task. The second requirement is to obtain
numerical ratings for each scale that reflect the magnitude of that factor in a given task. The overall workload score for each subject is computed by multiplying each rating by the weight given to that factor.

In Figure 53 we present the weighted rating for the TLX scores for each subject, under both IT conditions. The TLX scores of the first three subjects were lower for Stages than for FUI, meaning that Stages required less cognitive overhead than FUI. The fourth subject was the only subject who assigned a higher TLX score to Stages than to FUI. Subject's four subjective TLX ratings are not correlated with the facts reflected by our dependent measures, which show this subject performed much better with Stages than FUI: response time was 321 s and 399 s, respectively, the number of mouse clicks was 34 and 97, respectively, and the accuracy was identical under both conditions.

![NASA Taskload Index](image)

**Figure 53: TLX ratings for the four subjects, under both IT conditions**

### 6.10 Subjective comments

Upon completion of both sets of trials, our subjects were asked to freely comment on their experience with the two interaction techniques. Overall, the general consensus was that Stages interface "seems simple enough," involves "less work," "seemed to go faster," "icons [thumbnail controls] were easy to understand" and was "easier to use" than FUI. When asked to explain their preference for Stages, our subjects remarked that it was "much easier with stages because it takes less clicks to get where you want to go and you do not need to keep tabs of which study you’re on," or Stages "is less work because
you do not need to worry about clicking on the image before clicking on the controls to change which image is being displayed.”

The criticism for FUI was due mainly to the following major drawbacks of the interface:

- More cognitive overload was required to place the correct image pair in the two viewports,
- More manipulation (extra point-and-click actions) was required and it was more error prone,
  - "when you want to look at the second study, it takes 4 clicks instead of the 1 you need to make in the Stages interface"
  - "Confusing – which one you selected"
  - "Do not like the way it inverted by accident [first study with second study]"
  - "I always forgot to click on image before clicking on the button, so sometimes the image went in the wrong place"
- it provided no feedback on what images you had currently displayed on screen
  - "not confident using this interface,"
  - "found the UI confusing at first, not knowing if the icon was current in the view window"
  - "I was looking for some kind of user prompt to verify that the correct image was in the correct viewport"
  - "Annoying, do not know where you are in terms of time [first study versus second study]"
- it was not built to support the user’s mental modal, to target the completion of the user’s task
  - "did not like the interface"
  - "in the hospital, the images would be in sets, more like Stages interface"

6.11 Discussion and lessons learned from the pilot

Since only 4 subjects participated in our pilot study, no statistical tests were performed on our dependent measures between the two conditions (Stages vs FUI).

In order to ease-up our task of assessing the accuracy of interpretation using video-analysis of the experiment, we identified each trial with a number visible on both the ‘Start trial’ and ‘Stop trial’ buttons.
In the pilot experiment, neither the images, nor the controls had any labels to help the users identify what images are displayed on screen. We made this decision on purpose: since the state of the screen (as reflected by the content of the two viewports) could not be identified by using visual cues, our subjects had to keep track of the sequence of steps performed from the initial state (immediately after clicking on ‘Start Trial’, the two images from the first study were displayed on screen). And this is what our subjects did, as we can find out from their comments: “a caption underneath the navigational buttons would help to situate the user; however, I have the same order of clicking for each trial so I rely on this more than the interface to know which study I’m looking at.” An immediate benefit with Stages was that a single click on any of the four controls was enough to put the system in any one of the four possible states a subject may find useful for the completion of the task (let us not forget that stages were used to model a state in the user’s quest for solving a particular task). With FUI, four point-and-click actions were required to produce a similar result, and the order of the clicks was essential. Because of the corresponding time penalty associated with these interactions (required for bringing the screen layout to an intended state), our subjects had to work harder with FUI to keep track of the state transitions produced by their actions. Here are some of the comments of our subjects relevant to this subject:

- “It is difficult to know which one is study 1 and which one is study 2 – they look too similar in the icons.”
- “I would like to have differentiation between the timing between first and second study”

However, in a real radiological interpretation, the radiological images are labelled with the patient demographic information and particularly with the date and time when these images were acquired. For this reason we changed the presentation of stimuli in the main experiment to include a stamp (either ‘1’ or ‘2’) of the study they belong to for all our images and all the thumbnail-size controls.

By video analysis, we can say our subjects’ performance got better over time, independent of the type of interface and more reflective of overall comfort with the task. One subject noted “I got more familiar with the task as time went on.” Since we wanted to minimize the learning effects of the ‘breaking-in’ period needed for our subjects to
become comfortable using our system, we increased the number of training cases for the main experiment to five trials per each IT condition from 3 in the pilot.

Because of the perceptual component of our stimuli, we decided to control the complexity of our artificial images by creating three classes of complexity, based on the difficulty of detecting a target (number of distractors, eccentricity of the targets, conspicuity of the target etc.) When varying the difficulty of a trial, we had in mind an average trial completion time of 15-30 s, which is the approximate time required for the interpretation of a chest X-ray exam.

When dictating, radiologists are rarely annotating or marking the image to help identifying a lesion next time when that study is opened. The reason is anatomical landmarks are used in the radiological report to specify the location of a lesion. However, for teaching/presentation purposes, or when images are attached to the radiological report sent to the referring physician, the radiologist may highlight a target by circling it, or by drawing an arrow that points to the lesion or the area of interest. So we asked our subjects to use similar ‘landmarks’ in the reporting of their ‘diagnoses’. Consequently, our subjects had to identify the targets by verbally specifying the shape containing the target, if the target was enclosed, or by saying the target is “free floating” for non-enclosed targets.

However, naming the various shapes enclosing the target induced variation in their description. For example, the terms ‘hexagon’ and ‘stop sign’ were used interchangeably. We also noted a learning effect until the vocabulary associated with each shape was built by each subjects, and also additional delays with counting the vertices of a polygon, as we used hexagons, heptagons, and octagons. Since we noticed that most of our subjects used the mouse cursor to point to the target anyway, we decided for the main experiment to ask the subjects to identify the target solely by pointing with the mouse cursor. Thus we traded uniformity in identification of the target, but we increased the difficulty of the task with an additional mouse pointing operation. Since the workstation controls were located at the top of the screen, the time penalty with target identification would depend on the location of the target.
In summary, the pilot experiment was used to provide us with preliminary information on the distribution of eye-gaze over the screen layout, and helped us fine-tune the experimental conditions with the following improvements from the pilot setup:

- Labelled the secondary controls ‘Start Trial’ and ‘Stop Trial’ with the trial number, so we can recognize the current trial when watching the recording of a session.

- Labelled each of the 4 images of a trial with either ‘1’ or ‘2’, representing their corresponding study number.

- Labelled the controls thumbnails to fully identify the image(s) the thumbnail stands for.

- Asked our subjects to identify the targets by pointing to them, rather than describing the polygon containing the target.

- Create stimuli with three levels of complexity: low, medium, and high.

- Distribute the stimuli evenly across the following factors:
  - location: top/central/bottom, left/right
  - outcome: 00, 01, 10, 11, and 11 changed
  - complexity: low, medium, high
  - shapes: interchanged 3 shapes with other 3 shapes
7 User study

7.1 Experiment overview

Our experiment is based on an abstraction of the medical tasks of radiologic detection, diagnosis and quantification of abnormality extent. In order to use as subjects graduate students with no radiological background instead of radiologists, we designed artificial stimuli to be used instead of standard chest radiographs. As for the Pilot Study presented in the previous section, we are testing the performance of subjects engaged in a radiology look-alike conjunctive visual search under two different interaction techniques: Stages and FUI. For a detailed presentation of the experimental conditions, refer to Section 6.6 from the Pilot Study chapter. This user study has been approved by the Simon Fraser University Office of Research Ethics (see Appendix H: Approvals from the Office of Research Ethics).

7.2 User Study - Experimental design

7.2.1 Hypothesis

Our hypothesis was that using workflow oriented stages streamlines the radiological interpretation task, which leads to shorter completion times. For a detailed presentation of the hypothesis, refer to Section 6.3 from the Pilot Study chapter.
7.2.2 Task description

A detailed presentation of the task description was presented in Section 6.6 from the Pilot Study chapter. The subjects were required to find a target on a grey, complex background. A search using a conjunction of features was required for the discrimination of targets from distractors. Full details of what constitutes a target, and what constitutes a distractor, are given in a 15-slide ‘Training script’, given in Appendix D: Training script for the user study.

7.2.3 Procedure

The procedure for the user study reflects the improvements resulting from the pilot study, as presented in Section 6.11. For this within-subjects experiment we used 10 male and 10 female subjects. Each subject performed two consecutive blocks of 15 trials, one block for each interaction technique. The same 30 trials were performed by each subject. Inside each block, the order in which the trials were presented to the subjects was randomized. We counterbalanced for the interaction techniques, the subject’s gender, and the two sets of stimuli, as shown in Table 16.

Table 16: Counterbalancing for the interaction technique and the two sets of 15 stimuli each

<table>
<thead>
<tr>
<th>Subjects</th>
<th>First block</th>
<th>Second block</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 males + 3 females</td>
<td>Stages with Set$_1$</td>
<td>Free UI with Set$_2$</td>
</tr>
<tr>
<td>3 males + 2 females</td>
<td>Stages with Set$_2$</td>
<td>Free UI with Set$_1$</td>
</tr>
<tr>
<td>3 males + 2 females</td>
<td>Free UI with Set$_1$</td>
<td>Stages with Set$_2$</td>
</tr>
<tr>
<td>2 males + 3 females</td>
<td>Free UI with Set$_2$</td>
<td>Stages with Set$_1$</td>
</tr>
</tbody>
</table>

The procedure used for training of our subjects was formalized from the verbal one used for the pilot study. Appendix D: Training script for the user study contains the slide show we designed for training. Using PowerPoint the user can control the speed for the presentation of information.

After learning about their task, the subjects were introduced to the application used during our experiment. The layout of the screen consisted of a left viewport, a right
viewport, the controls used for image selection, and the controls used to start/stop the current trial, as illustrated in Figure 54.

For the completion of each trial, our subjects had to perform the following steps:

1. Before each trial, an empty screen was displayed. The only button enable was the ‘Start Trial’. A screen shot corresponding to this step using Stages is shown in Figure 54.

Figure 54: An empty screen shows no stimuli. The only enabled button is ‘Start Trial’.

2. When a subject clicked on ‘Start Trial’, a timer started to record the response time. The ‘Start Trial’ button became disabled. The ‘Stop Trial’ button and all the thumbnail buttons became enabled. The two images corresponding to the first study were automatically displayed in the two viewports as shown in Figure 42. The subject had to find if a target was present in this first study. If
a target was found, the subject had to say "First study: here is the target" and point at the target with the mouse cursor. If no target was found, the subject had to say "First study: no target."

3. After the first study was interpreted, the subject had to examine the second study. The two images corresponding to the second study had to be displayed in the two viewports. This was achieved by simply clicking on the second thumbnail, if Stages was used. With FUI, four clicks were required to produce the same result, as shown in Figure 43: first the subject had to click on the left viewport, and then on the third thumbnail to bring down the PA image from study two. Then, the subject had to click on the right viewport, and then on the fourth thumbnail to bring the LAT image from study two. The subject had to search for a target in study 2. If a target was found, the subject had to say "Second study: here is the target" and point at the target with the mouse cursor. If no target was found, the subject had to say "Second study: no target."

4. In this step, an image from each study was displayed on screen in order to assess the evolution of the target’s size. It corresponded to Stage 3 (shown in Figure 55) or Stage 4 from Table 13. This step is optional, and only makes sense if a target is found in both studies. However, if the subject was confident in assessing the target’s evolution, such was the case with many trials where the target size remains the same, this step was skipped by the subject in order to reduce the response time.
Figure 55: This layout corresponds to Stage 3 of a medium complexity trial, and shows the PA images from study one and two. Each image has 4 distractors and 3 wild cards.

5. The subject had to click on 'Stop Trial' after finishing the trial interpretation. The timer associated with the trial was stopped automatically, and the empty screen was shown again. The 'Stop Trial' button and all the thumbnail controls became disabled. The subject could not go back to see the images from the trial that was just completed. As a difference from the pilot, we introduced a UI to record the user subjective ratings for the perceived response time and perceived complexity. The radio buttons used to collect these subjective ratings were the only controls enabled, as shown in Figure 56. Since these subjective ratings were collected at the end of each trial, they did not affect the response time. After the subject selects one option from the response time, and one option from the trial complexity, the 'Start trial' button
became enabled. The system was now back to step 1, ready to start another trial.

Figure 56: An empty screen displayed after the completion of a trial. The radio buttons used to collect subjective ratings are the only controls enabled.

The steps described above are always required in order to properly and optimally complete a trial. However, additional steps can be introduced by a user, which derail from this optimal interpretation strategy. For example, a subject may choose to revisit one of the previously completed steps. It is also possible for a subject to accidentally stop a trial immediately after interpreting the first study. Such an event will generate an incorrect interpretation, as the second study was completely missed.

7.2.4 The stimuli

We wanted to minimize the learning effects by increasing the number of training cases. So in addition to the training sessions presented in PowerPoint, we used another 5
warm-up cases for each interaction technique, to give our subjects the possibility to become comfortable with the task, and to reduce the learning effect. We also wanted each trial to last between 15 s and 30 s, which is the approximate time required for the interpretation of a chest X-ray exam.

7.2.4.1 Target distribution

The target distribution (TD) reflects the presence of a target in the two studies from a trial. Therefore, there are five classes of trials:

- trials with no target in either study (00),
- trials with no target in the first study, and a target in the second study (01),
- trials with a target in the first study, and no target in the second study (10),
- trials with a target in both studies, and the target size remained the same size (11),
- trials with a target in both studies, and the target shrunk or grew in size (11 changed).

The trial outcome can influence a subject’s performance by affecting:

- the visual search strategy: since there is at most one target, once a target is found, which may happen immediately after the search started for low-complexity trials, the search is terminated. Also, if a target is found in the first study, no exhaustive search is required on the second study.
- the cognitive workload and the response time: if a target is found in both studies, an extra step may be required involving an inter-study image comparison to assess the evolution of the target size.

A second classifications scheme that we considered at first was:

- no target in any study,
- only one study has the target, so there is no need to comment on target evolution in size,
- both studies had a target, with or without changes in the target’s size.

A third classification scheme we considered was based on outcome and optimal search strategy involved:

- trials where the first study has a target. This means no exhaustive search on second study is required, so image complexity on the second study should have little impact on the subject’s performance.
- trials where the first study has no target.
7.2.4.2 Trial or screen complexity (TC)

Because of the perceptual component of our stimuli, we wanted to create three classes of stimuli, based on the difficulty of detecting a target. The idea behind rating the complexity of the stimuli was that the impact (caused by the two interaction techniques) on our subjects' performance would be increased under heavy cognitive load. Consequently, we generated three levels of trial complexity (low, medium, and high) by changing the following three factors: target conspicuity, configuration similarity, and location, as presented in Table 17. These factors, and the values associated with a particular trial complexity, were generated by analysis of the results from the pilot study: response time, accuracy, and number of mouse clicks. While our initial approach was to give different weights to these factors, in the end we decided to give them the same importance, in order to simplify the process of classifying the stimuli into classes of trial complexity.

Table 17: The three levels of trial complexity were generated by varying the target conspicuity, the configuration similarity, and the target location.

<table>
<thead>
<tr>
<th>Trial complexity</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target conspicuity</td>
<td>Target appears in a frame, and has high contrast.</td>
<td>Either the target appears in a frame and has low contrast, or the target appears with no frame and has a high contrast.</td>
<td>Target appears with no frame and low contrast.</td>
</tr>
<tr>
<td>Target location</td>
<td>Central.</td>
<td>On extremity.</td>
<td>No target.</td>
</tr>
<tr>
<td>Configuration similarity</td>
<td>Low.</td>
<td>Medium.</td>
<td>High.</td>
</tr>
</tbody>
</table>

The target conspicuity was reflected by the target enclosure and the target contrast relative to background. Sample targets of low, medium and high conspicuity are shown in the following image pairs: Figure 57 and Figure 58, Figure 59 and Figure 60, and Figure 61 and Figure 62 respectively.

The pilot study revealed the subjects prefer to scan the first image from left to right, top to bottom. Sometimes a quick inspection in the centre of the image is done.
before proceeding with the systematic search. So we distributed our target location evenly across the screen, with central-located targets in low-complexity trials, and targets located on the extremity for medium-complexity trials.

The configuration similarity was controlled using the *distractor-distractor similarity* and the *number of shapes*, as described in Table 18.

Table 18: Rating system for configuration similarity

<table>
<thead>
<tr>
<th>Trial complexity</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distractor-distractor similarity</td>
<td>High (heart vs. disc)</td>
<td>Medium (discs divided incorrectly)</td>
<td>Low (octagon vs. disc)</td>
</tr>
<tr>
<td>Number of shapes</td>
<td>2 wild cards and 3 distractors</td>
<td>3 wild cards and 4 distractors</td>
<td>4 wild cards and 5 distractors</td>
</tr>
<tr>
<td>Configuration similarity mark</td>
<td>0 (high)</td>
<td>1 (medium)</td>
<td>2 (low)</td>
</tr>
</tbody>
</table>

**7.2.4.3 Examples of stimuli**

Figure 57 and Figure 58 show two examples of low complexity stimuli, with no target present. Each figure includes 2 wild cards and 3 distractors. Hearts were used as distractors instead of discs.

![Figure 57: Study 1, PA view](image1)

![Figure 58: Study 1, LAT view](image2)

Figure 59 and Figure 60 show two examples of medium complexity stimuli, with the target present in the octagon. Each figure includes 3 wild cards and 4 distractors. Incorrectly divided discs were used as distractors.
Figure 59: Study 1, PA view

Figure 60: Study 1, LAT view

Figure 61 and Figure 62 show high complexity stimuli. The target is in the circle. Each figure includes 4 wild cards and 5 distractors. Octagons were used as distractors.

7.2.5 Independent variables

Our experiment has three independent variables: the interaction technique (IT), the target distribution (TD), and the trial complexity (TC). There are two conditions for IT, Stages and FUI respectively. Based on TD, we can distinguish the following conditions: 00, 01, 10, 11 same, 11 changed. In terms of TC, our trials had low, medium, or high complexity. Consequently, our experimental design had $2 \times 5 \times 3 = 30$ distinct cases. We associated one trial per case (i.e. a set of 2 studies, each study has 2 images), for a total of 30 trials. The stimuli for the 30 trials were grouped into two disjoint sets (Set$_1$ and Set$_2$) of 15 trials, one for each interaction technique. For each set, the order in which the 15 trials were presented to the subjects was randomized.
7.2.6 Dependent measures

Response time (RT): individual trial completion times were recorded in a user-specific log file. RT was measured from the moment the subject clicked on 'Start trial' to have the stimuli displayed, until the subject clicked on the 'Stop trial' and the stimuli were hidden. There were 20 subjects x 30 trials = 600 RT measurements.

Interpretation errors: the interpretation accuracy of each trial was assessed by video analysis. The video footage captured both the screen content and the subject’s verbal interpretation. There were 20 subjects x 30 trials = 600 trials for which interpretation errors were measured.

Mouse clicks per trial: used to measure the number of user interaction steps required for the interpretation of each trial. There were 20 subjects x 30 trials = 600 for which the number of mouse clicks were counted.

User cognitive overload - NASA Task Load Index (TLX). There were 20 subjects x 2 IT = 40 TLX measurements.

Perceived ease of use/comfort – we used the System Usability Scale (Q10) and our own questionnaire (Q19). Details on these questionnaires are presented in Appendix E: System Usability Scale (SUS) and Appendix F: Second usability questionnaire (Q19), respectively. There were 20 subjects x 2 IT = 40 Q10 and respectively Q19 measurements.

7.3 Results

7.3.1 Response Time (RT)

The following abbreviations are used in this section:

- IT: Interaction Technique (Stages or FUI)
- SET: which of the two sets (each set had 15 trials) was used first by the subject
- METHOD: which one of the IT was used first by the subject

The RT data is summarized in Figure 63.
In order to approximate the normal distribution, we used the Normal Q-Q Plot to compare the quantiles of the RT distribution against the quantiles of the normal distribution. The RT distribution did not closely match the normal distribution (the points did not cluster around a straight line). We applied the natural logarithm to the RT data, and the Normal Q-Q Plot showed a normally distributed data. We then performed a General Linear Model (ANOVA) on the logarithm of RT, which showed the interaction technique had a significant effect on RT ($p<0.001$). The average response time was 17.0 s and 19.7 s for Stages and FUI, respectively. Since we used the natural logarithm of the RT for our statistical analysis, we will also mention the RT corresponding to the average $\ln(RT)$ values under both IT conditions: 2.77 for Stages, and 2.93 for FUI, which corresponds to 15.8 s for Stages, and 18.4 s for FUI. The average saving in RT due to the use of stages was 14%, independent of the reference used, RT or $\ln(RT)$.

Figure 64 shows the RT plotted against the trials outcome. The trial's Outcome had also a significant effect ($p<0.001$) on RT. However, Complexity, shown in Figure 65, had no significant effect ($p=0.319$) on RT.
We also looked for differential learning effects, and we used SET and METHOD to reflect the first set (and respectively method) used during the experiment. The Test of Between-Subjects Effects showed METHOD to have a marginal effect (p=0.053) on the RT. Gender had no significant effect (p=0.435). SET had no significant effect either (p=0.20), so we successfully distributed the stimuli in the two sets to avoid effects on the experimental results.

We recorded a significant effect for IT * METHOD (p<0.001), as shown in Figure 66. For the 10 subjects that started with FUI, there was a significant improvement in performance when they performed under stages, as illustrated by the reduction in the average RT. However, for the 10 subjects that started with Stages, their performance did
not improve in the second part of the experiment, when they performed with FUI. So only with Stages were our subjects able to reduce their RT as they became more familiar with the task. Consequently, we believe Stages is a better interaction technique, as it allows the users to optimize their workflow and in to increase their productivity as they become accustomed with the main task, the visual search for targets.

![Estimated Marginal Means of LN(TIME)](image)

Figure 66: IT* METHOD had a significant effect on RT (p<0.001).

### 7.3.2 Mouse clicks (MC)

The average number of mouse clicks per trial was 2.29 for Stages, and 6.26 for FUI. The standard deviation was 0.58 for Stages, and 1.18 for FUI.

As expected, the IT had a significant effect (p<0.001, df = 19) on the number of clicks required to complete a block of 15 trials, with MC_{avg} = 34 for Stages, and MC_{avg} = 94 for FUI. A full analysis using a paired T-test is shown below in Table 19.
### Table 19: Paired Samples Statistics for Mouse Clicks

<table>
<thead>
<tr>
<th></th>
<th>Mean number of clicks / block</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stages</td>
<td>34.40</td>
<td>20</td>
<td>3.912</td>
<td>.875</td>
</tr>
<tr>
<td>FUI</td>
<td>93.85</td>
<td>20</td>
<td>7.307</td>
<td>1.634</td>
</tr>
</tbody>
</table>

#### 7.3.3 Interpretation errors

Our subjects were instructed to be as accurate as possible in their diagnosis, as their primary task. Completing each trial in the shortest possible time interval was a secondary requirement. This is the reason our hypothesis made no references to the distribution of errors between the two interaction techniques: we traded time for accuracy. However, 27 errors were made when FUI was used, and only 17 errors were made with Stages. There were four types of errors, as shown in Table 20, and their number is presented in the format [Stages, FUI]:

- search errors in study one and two (S1+S2), such as missing a target, or taking a distractor as a target [17, 29];
- usability errors, such as making the diagnosis by looking at the wrong pair of images [0, 9];
- evolution errors, meaning the target's evolution in size was incorrectly assessed [0, 2].

Table 20 shows the distribution of the 4 types of errors based on the interaction technique. Table 22 and Table 23 show how errors were distributed for male and female subjects, for the two interaction techniques. Figure 67 shows the error distribution for each of the five different outcomes. Two male and two female subjects made no interpretation errors for both FUI and Stages. For Stages only, three male and three female subjects made no interpretation errors. For FUI only, two male and four female subjects made no interpretation errors. We elaborate on these results in the Discussion section.
Table 20: Interpretation errors by type

<table>
<thead>
<tr>
<th></th>
<th>S1+S2</th>
<th>Wrong image</th>
<th>Evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stages</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FUI</td>
<td>29</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 21: Error distribution for male subjects. Two subjects made no interpretation errors.

<table>
<thead>
<tr>
<th></th>
<th>FUI</th>
<th>Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1+S2</td>
<td>Wrong image</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>
Table 22: Error distribution for the 10 female subjects. Two subjects made no interpretation errors.

<table>
<thead>
<tr>
<th>FUI</th>
<th>Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1+S2</td>
<td></td>
</tr>
<tr>
<td>Wrong image</td>
<td>1</td>
</tr>
<tr>
<td>Evolution</td>
<td>2</td>
</tr>
<tr>
<td>S1+S2</td>
<td>10</td>
</tr>
</tbody>
</table>

Total number of errors per outcome condition

![Bar chart showing number of errors per outcome condition](image)

Figure 67: Number of errors per outcome condition
7.3.4 Subjective ratings

Only the scores for the two user satisfaction questionnaires (Q10 and Q19) showed significant preference for Stages.

7.3.4.1 NASA Taskload Index (TLX)

The average TLX scores were 52.6 and 56.4 for Stages and FUI, respectively. There was no significant difference ($p=0.106$) between the two IT.

7.3.4.2 System Usability Scale (Q10)

The average Q10 scores were 82 and 74 for Stages and FUI, respectively. The standard deviation was 11 for Stages, and 17 for FUI. A significant difference ($p=0.03$) was recorded, as subjects gave a higher usability rating to Stages than to FUI in this standardized usability questionnaire. More information on this usability questionnaire is provided in Appendix E: System Usability Scale (SUS).

7.3.4.3 Usability questionnaire (Q19)

The average Q19 usability scores were 5.85 and 5.4 for Stages and FUI, respectively. A significant difference ($p=0.03$) was recorded, as the subjects gave a higher usability rating to Stages than to FUI on a scale from 1 to 7. More information on this usability questionnaire is provided in Appendix F: Second usability questionnaire (Q19).

7.3.5 Subjective comments

Most subjects preferred to use Stages, and described it as 'straight-forward', 'normal interaction', 'easier to use', 'easier to learn and operate', 'less work'. FUI was criticized ('annoying', 'mental and physical workload') for the extra clicks required, the extra flexibility not being useful for the given task.

When asked which IT they would prefer if they had to do the same task again, only subject 10 and subject 20 chose FUI. However, their performance was much better for Stages as shown in Table 23.
Table 23: The performance for the 2 subjects that preferred FUI

<table>
<thead>
<tr>
<th>Subject number</th>
<th>Total number of errors</th>
<th>Average RT (s)</th>
<th>Total number of clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stages</td>
<td>FUI</td>
<td>Stages</td>
</tr>
<tr>
<td>10 (female)</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>20 (male)</td>
<td>1</td>
<td>6</td>
<td>17</td>
</tr>
</tbody>
</table>

7.3.6 Discussion

The interpretation task was purposely designed to put a high load on a subject’s attention and working memory:

- the target definition required a conjunctive visual search, and thus remembering the characteristics of a target (two discs AND same diameter AND half greyed AND split in the same direction)
- the use of a ‘wild card’ implied remembering the location on the first image, and finding the corresponding location on the second image of the same study in order to ‘solve’ the wild card
- when a target was detected in the first study, its location had to be remembered in order to speed up the interpretation of the second study
- when a target was present on both studies, the evolution of the target had to be assessed, which implied remembering the size of the target from the first study in order to avoid the extra time/clicks involved with creating a new screen layout (one image from study 1, and one image from study 2).

Our experimental results showed a significant difference in the RT for the two IT, with a 14% increase in the average RT, from 16.9 s for Stages to 19.7 s for FUI. Only 4 subjects had a shorter RTavg for FUI than for Stages, and they all started with Stages. Since all 4 subjects expressed their preference for the use of stages, our educated guess is that these differences in performance were due to the learning effect14. We believe the difference in RT under the two IT is not only due to the extra 4 clicks/trial, but also to the cognitive workload, and to the fact some of the information presented above was

14 The experiment took about 1h for each subject. We chose not to use more trials for training to avoid effects such as fatigue and boredom.
‘flushed’ from the working memory in order to ‘load’ the 4 point-and-click steps required in FUI.

Here are some comments from different subjects supporting our hypothesis:

- “remembering the position of the target can be more difficult with FUI (more time elapsed, more interactions)”
- “extra clicks are a distraction to remembering”
- [liked Stages because] “I did not have to keep track of which image is where, to be responsible of the screen layout”
- “too much flexibility in FUI, especially when trying to determine which images to bring down for inter-study target evolution”
- “less distracted by the interaction in Stages”
- “FUI requires more clicks, and it takes more time to arrange images, and short term memory is very important in this task.”

In our experiment, FUI caused 2 errors due to incorrect interpretation of the target’s evolution (no such error was produced with Stages). FUI also produced 9 errors due to a fault-prone IT: the subjects meant to analyze the image from the second study, but they were not examining the appropriate images. Such error is not technically possible when using Stages. It is worth noting this effect occurred despite the fact each image had a clear label on which study it belonged to, and also that subjects can compare two instances of the same image despite the visual cues offered by the PA / LAT background.

The most difficult outcome condition was probably when a target, present in both studies, changed its size. We base this belief on the results reflected by dependent measures such as RT and number of errors. Under this outcome condition the differences between Stages and FUI are most visible: the RT was 14.8 s and 20.8 s, and the errors were 1 and 6, for Stages and FUI respectively. This leads us to believe the benefits of a good IT, such as Stages, are more likely to be visible under heavy user overload.

This is the typical situation of a highly repetitive task that has stringent requirements of accuracy and speed. The user thinks (s)he learned the sequence of interactions: 4 point-and-click steps in our case: click on the left viewport, click on PA\textsubscript{2} thumbnail, click on the right viewport, and click on LAT\textsubscript{2} thumbnail (relies on this
sequence to provide the appropriate data for interpretation). This is a major danger in highly repetitive tasks: even when the information of image label / study acquisition time is clearly displayed on the screen to prevent such errors, this information becomes irrelevant to the subject as they rely on an interaction pattern to produce the correct images on the screen. Bringing the second study for interpretation using FUI is more error prone not only because more steps are involved, but also because the order of the steps is critical.

However, our subjects did not have enough practice to become expert users, and the 4-steps interaction sequence did not migrate into their ‘muscle’ memory. This is also the case with many radiologists, they do not spend enough time with one workstation to advance from the level of beginner/intermediate user to the level of expert/power user. And the reason is they either read studies in various offices with different software products, or because the interaction in new software releases is not consistent with the previous version. There are known situations of radiologists reporting an abnormality on one side, when it should have been reported on the other side. The problem is usually caused by the fact that most radiologists expect to see the right side of a patient displayed on the left side of the screen, and they sometime rely on the software to flip the image accordingly, and do not check the patient’s orientation using the markers displayed on the images.

User satisfaction was considerably higher with Stages than with FUI. Eighteen of the twenty subjects said they would use Stages and not FUI if they had to do another block of trials. The two usability questionnaires also showed significant difference in the ease of use, with Stages being the preferred IT.

It is interesting to note the ‘workarounds’ some people came up with to improve their performance when using FUI. One subject “liked FUI better, because you can compare after one click if the target changed size” (when a target was present in the first study). Another subject, when a target was visible in the first study, would:

- “Look for the change in size” [of the target, after replacing one image from the first study with an image from the second study], and thus avoid the extra
time/clicks involved with setting up the screen for assessing the target's evolution [if a target was to be found in study two];

- "See if the two discs swapped [top-bottom], and looked if they swapped again" rather than remembering the position of the wild card from the first image.

In conclusion, several dependent measures, such as RT, number of clicks, number and type of errors, and user subjective ratings show that Stages produced better results. These results lead us to believe that vendors should work closely with the radiologists to build staged hanging protocols for various radiological scenarios, such as 'abdominal CT', 'MRI of the spine', and 'X-ray of the chest'.

### 7.3.7 Improving our interaction technique

Our goal was not to come up with the perfect IT for the task we devised. The fact our subjects had to point to the target with the mouse was probably distracting, especially in Stages, as subjects had to move the mouse from the control area (situated in the top-left corner) down towards the center of the screen. So it came as no surprise to us when some of our subjects made valid suggestions on how the IT can be improved. One subject suggested eliminating the thumbnail button corresponding to Stage 4 (LAT$_1$ <-> LAT$_2$), since it provides redundant functionality (inter-study image comparison for assessing the target's evolution) with Stage 3 (PA$_1$ <-> PA$_2$). A couple of subjects suggested using keyboard shortcuts instead of our thumbnail controls. This is exactly what Dr. Steven Horii, a radiologist from U. Pennsylvania wrote referring to mapping preset values for contrast/brightness to keyboard shortcuts (Horii 2003a):

...availability of appropriate controls. For example, have the window width and level presets and any manipulation tools (zoom, roam, etc.) available quickly as needed. Not all the vendors can do this (Horii 2003b).

### 7.4 Feedback from radiologists on the user study

We started validating our results with physician subjects. At the time this thesis was distributed, experiments with other radiologist subjects were taking place, both with and without the use of eye-trackers. Our preliminary data suggests physicians perform a
more thorough search, as a habit of their profession, which leads to longer response times.

7.4.1 Medical Doctor

Subject 1 is an MD with an academic appointment in the radiology department of a University hospital. He noted that “you put lot of work in designing the UI and the stimuli,” referring to the difficulty of registering two images due to slightly shifted objects and the use of various shades, which “stress the task and highlights the benefits of a good UI.” He thought the use of the wild-card was “diabolically clever,” and designing the wild-card to “flip” on the second image gave our search task the right level of difficulty, making it “more like the radiologist’s task.” He also remarked that having the objects going outside/close to the PA/LAT background borders, slightly shifting in the target’s location, making some objects disappear on the second image, and varying the location/size/enclosure/grey level of the (potential) targets provided a “challenge similar to the radiologist’s task, and proves understanding of the noise embedded in the structure of radiological images.”

This subject performed the first block of trials with FUI, and the second block of trials with Stages. When comparing the two interaction techniques, this subject noted the FUI induced a considerable level of fatigue, making the task surprisingly difficult. However, he was considerably less fatigued after using Stages, due to the reduced cognitive overload associated with UI manipulation. After trial 10, he became comfortable with the task, the interpretation rate started to pick up, he “started letting go, as I did not have to worry anymore about eye-hand coordination, I did not have to match the images from each individual study anymore, and I did not have to compare for size changes.” Under the Stages condition, this subject performed inter-study registration (in order to asses if the target changed its size) only for the trials where a target was present in both studies, and the target had the same size. For all the other trials, this subject

\[15^{*}\] He didn’t have to concentrate on arranging the image-pairs for comparison. It was also easier to perform the comparison for assessing the target’s change is size, so he didn’t have to concentrate anymore on the size change like in FUI.
clicked only two times on the thumbnail controls (i.e. he performed no inter-study comparison for trials where a target was visible in both studies, and the target changed size).

This subject’s average response time was 37 s for FUI and 21 s for Stages, while the average for our experiment was 20 s and 17 s. In terms of mouse clicks, this subject totaled 113 clicks for FUI and 33 clicks for Stages, while our study average was 94 clicks for FUI and 34 clicks for Stages. This subject’s comments after FUI was that he got confused with the FUI two times: it seemed to him he “advanced to displaying both images of study 2, without me acknowledging it.” For those two trials when he got confused, he had a big response time (108 s and 91 s respectively), and he performed many clicks (17 and 18 respectively). The irony is those two trials were the only ones he incorrectly diagnosed! If we exclude these two ‘problematic’ trials from the FUI results, his average RT for FUI becomes 27 s, and he clicked exactly 6 times on every FUI trial, for a total of 90 clicks.

Our educated guess is that FUI does not provide a simple way to ‘reset’ the system to a known state. Consequently, error recovery is more difficult to achieve if FUI is employed rather than Stages.

7.4.2 Radiologist

The second subject is a practicing radiologist with experience on different PACS workstations. He first performed under Stages, and then under FUI.

This subject’s average response time was 27 s for Stages and 25 s for FUI. In terms of mouse clicks, this subject totaled 96 clicks for FUI and 36 clicks for Stages. Four diagnostic errors were generated with Stages, all of them in the second study. There were no interpretation errors generated with FUI.

After performing a task analysis of the videotape of the session, we think there are two major factors that contributed to a better performance for this subject under FUI. First, there was a significant learning effect. For example, the subject spoke aloud in the first few trials, which increased his RT. Also for the first four trials, there was a delay of
about five seconds from completing the interpretation until the ‘Stop Trial’ button was clicked. On trial 8 with Stages, no target was found in the first study, so the subject decided to double check his diagnosis by performing a second full search on this study. The subject asked the experiment supervisor if a target is always present in a target. Then he clicked on ‘Stop Trial’ without inspecting the second study. Thus a diagnosis error was recorded, and the response time of 27 s only involved the search of the first study from a medium-complexity trial.

7.4.3 About the performance of our physicians

In my first two years of high school, I would sometime find a solution to a mathematical problem ahead of my professor, who had a very good, and well merited, reputation. Later on I realized this ‘paradox’ was only partly due to my math skills. Due to my limited knowledge in the field, I only had 1 or 2 approaches to that genre of math problems. However, my professor could think of many more ways of taking on that problem. So I was probably faster in finding a solution when my approach was right on target, and when my professors’ initial approach either failed or took longer to find a solution.

The same is probably true for my two M.D. subjects. For this experiment, they had to unlearn habits acquired during the many years of training as physicians. For example, during the interpretation, the two physicians would sometime refer to the target as a ‘lesion’. They were also more prone to providing additional, unnecessary information about the target’s evolution, such as the target moved, or that the split disc and the wild card switched positions. This is probably due to the fact physicians may simply be more careful/cautious/complete, and operate with a different criterion on the speed-accuracy tradeoff. Also, their expertise in recognizing anatomical abnormalities in radiological images did not transfer to artificial targets, or only transferred partially, at the “abstract level of task analysis, and not at the perceptual level,” as our radiologist subject noted.
8 Conclusions

8.1 Major contributions

8.1.1 Radiological interpretation: softcopy vs. hardcopy reading

We performed a questionnaire-base user study in November 1999, meant to identify the requirements for a next generation radiology workstation. This study involved eight radiologists who answered questions and provided comments on three complementary research topics, as described in Section 4.3.

First we asked our subjects to enumerate the advantages and the disadvantages for both softcopy and hardcopy reading. The main observation was that for digital radiology to succeed, one should avoid a one-to-one translation of the steps involved in the film-specific workflow. We identified the two major factors for productivity improvement through the use of PACS workstations, both related to the logistics surrounding the interpretation task: workflow reengineering and process automation. Therefore, we concluded that productivity improvement with softcopy reading is conditioned by the integration of process automation mechanisms such as study prefetching and hanging protocols (HPs).

Second we collected radiologist feedback on the use of HPs. Our goal was three fold: to asses the need for HPs, to validate our hypothesis towards a hierarchical structure of HPs, and to determine the average number of HPs required for the daily softcopy reading practice. The results indicated the high importance of automatic image
organization through HPs, with the potential effect of reducing the interpretation time by 10-20%. Our subjects estimated that 10 to 15 HPs would cover about 85% to 95% of the regular radiological examinations. The remaining 5-15% of examinations will be impractical to cover under HPs, since they represent either very rare examination types or exceptions from typical scanning protocols. These results depended on the modality, anatomical region, and the radiological procedure type. Consequently, the hierarchical structure of HPs that we proposed would provide a good mechanism for HP filtering and selection. Fewer than twenty HPs per modality will be convenient for the radiologist to generate, and for the radiology workstation to manage and select from. Another contribution of this study comes from highlighting the potential benefits of HPs, such as reduced interpretation time and less user fatigue. We also showed that only a small number of radiologist-specific hanging protocols are required, and that our proposed structure for the HPs based on modality, anatomical region, and radiological procedure type will provide a good mechanism for HPs filtering and selection. Our early conclusions on the use of HPs were recently confirmed by Dr. N. Strickland from the Hammersmith Hospital, a pioneer and an expert in radiology image display:

Whilst it is my belief that it is important that DDP [Default Display Protocols, another name for HPs] should be very flexible, I have come to the opinion, with experience, that the most important factors are:

* Personalized DDPs should be easy to set up by the user on his/her PACS.

* That the user should not be overwhelmed by a huge range of possible DDPs hardwired into the PACS by the Vendor.

* The majority of users do not bother customizing their own DDPs (and this includes radiologist users as well as non-radiological clinicians), but simply configure a suitable DDP on the fly...

There is no question that the vendors have not yet cracked the problem of HPs, and many have gone from one extreme to another, i.e. from having no or few possible HPs to having so many that the user is overwhelmed (Strickland 2002).

Third, we investigated the impact of the display devices on the radiologist’s workflow. Our results indicated that monitors with different properties would be required for different modalities. The monitors’ spatial resolution was very important for reading
projection radiography, less important for cross-sectional imaging, and of little importance for ultrasonography. Having multiple monitors was considered essential for reading CT and MRI, important for projection radiography, and of little importance for reading US.

These key results from 1999 on key functional requirements for softcopy interpretation only recently were incorporated in most of the current, successful PACS workstations. For example, the tools our study identified as the most commonly used, such as adjustment of window width and level, image zoom, and the stack mode, are now always available for immediate use. The speed of study loading and the optimal arrangement of images for interpretation is much improved nowadays by hanging protocols that also include the automatic retrieval and display of relevant prior examinations.

8.1.2 Task analysis for radiological softcopy interpretation

We took an interdisciplinary approach to identify key components of the radiologist's task, by using well established theories from different research domains.

By surveying the psychology literature, we identified research results from perception and cognition applicable to radiology softcopy reading. By studying perception we gained insight on the main causes for false negative errors. Using eye-fixation recordings, three causes were identified for false-negative errors: faulty visual search, faulty pattern recognition and faulty decision making (Kundel, Nodine et al. 1978; Krupinski and Roehrig 1999; Berbaum, Franken et al. 2000). These results were recently confirmed by one of our radiologist subjects, whose comments on the perceptual component of study interpretation was:

The biggest problem with radiology is just not perceiving if there is a lesion. And is not a question of analysis, it is a question of visual acuity, and perceiving that within among those shadows there is something that may represent a lesion... And the vast majority of errors are perception errors. It is not that it is a difficult lesion to analyse; it is that they did not even see there was even a shadow in there...
We used cognitive research to draw on the knowledge of structures and cognitive processes underlying radiologist performance. We ranked users of radiology workstations as experts, intermediates and novices. We recognized the existence of specific and predictable differences between novices and experts in terms of their perceptual search behaviours. So we hypothesised that novices can improve the speed and accuracy of their interpretation by using HP++ created by expert radiologists.

In Section 4.6.4 we also suggested extending the workstation functionality with aids to image perception using perceptually-based feedback. This can be achieved by recording the eye-gaze of the radiologist when searching for lesions and then have the workstation circle the areas associated with dwells longer than 1 second.

We were also able to apply general principles of good engineering design to our specific domain. Section 4.7. presents our workflow decomposition and task analysis, which produced the following characteristics of the radiologist’s task:

- It is time critical
- It involves a visual search using high-resolution displays
- It involves a complex workflow consisting of several states
- It has a highly contingent work flow
- It has a working memory filled by the task

We used principles from usability engineering to suggest solutions for the placement of controls in order to reduce the disruptions of the interpretation process, and especially interruptions of the visual search.

Finally, we used principles from software engineering to describe an iterative workstation design process based on rapid prototyping and frequent feedback from representative users, as outlined in Section 3.12. Our practical experience with the successful design and deployment of a state-of-the-art diagnostic radiology workstation provided invaluable insight for identifying the requirements of next generation workstations.

Designing Better User Interfaces for Radiology Workstations
8.1.3 Heads-up, streamlined radiology interpretation

In Section 5 we presented a comprehensive software approach for improving the radiology interpretation process by significantly reducing off-image gaze fixations, and by optimizing the highly repetitive tasks involved in image navigation and manipulation.

Based on our eye-gaze study described in Section 6, we classified the different levels of disruption of a radiologist's diagnostic process due to the interaction with the workstation, listed in order of increasing disruption:

- Visual attention on the features of the control rather than the features of the image under analysis.
- Eye movement from the image area of interest to locate a control and guide the hand in executing it.
- Cognitive load to recall, select, and execute controls.

We think there is great potential in lowering the cognitive workload of the radiologist with multimodal interfaces, while keeping an eye on the impact on the user due to unloading visual controls onto motor controls. PACS vendors should look carefully at the lessons learned from the videogames industry in terms of how to design ergonomic and unobtrusive controls and how to provide new hardware, such as joysticks and pedals, to facilitate the user interaction experience. For example, Ferraro Design's 'The Claw' provides a comfortable solution for a mini-keyboard configuration, having nine programmable buttons with up to five user-defined keystrokes each. The claw was designed for use in conjunction with the mouse, requiring no off-screen glance from the user. This kind of input device could successfully replace the standard keyboard, since the user's need for text input could be eliminated or reduced through predefined or 'canned' text messages and speech recognition software.

8.1.4 Pioneers for inexpensive usability studies

A 'typical' validation of our hypothesis would have required prohibitively expensive resources, including a fully-functional radiology workstation as the test bed, radiologists as subjects, and real-life radiological images as stimuli. With the goal of performing inexpensive usability studies related to radiology workstation design, we designed a new set of stimuli and adapted the experimental task in order to test our
hypothesis using a basic workstation and novice users as subjects, as described in Section 3.4. Since we are interested in the external validity and the generalizability of our results, we started to validate these results with local physicians. Initial comments from medical professionals suggest that our results from non-expert users can be transferred successfully to radiological softcopy interpretation tasks.

8.1.5 HP++ and the automation of routine tasks

In Section 5 we proposed extending the concept of HP to further support the radiologist’s reading task, by adding support for scenarios, stages, heterogeneous data, adaptation, virtual monitors, and type abstraction hierarchy.

When the radiologist selects a ‘main study’ for interpretation, the workstation looks for the best matching hanging protocol. The HP++ selection involves the analysis of a potpourri of information, corresponding to the patient, study, series or image level. With scenario based interpretation, a complex set of behind-the-scene actions are triggered, starting up by building the clinical context associated with the main study. Relevant radiological examinations are automatically pre-fetched, together with their corresponding radiology reports. This complex, heterogeneous data is grouped temporally in a sequence of stages, according to the mental paradigm of the physician.

8.1.6 Radiologist-centred interaction technique

The radiologist has highly repetitive interpretation task, with stringent requirements of accuracy, confidence, and speed. Accessing the controls of current radiology workstations produces considerable disruption of the visual search, which may lead to differences in the volume and type of information processed. We invented an interaction technique attuned to this primary task. In a radiology look-alike task, our user-centered interaction technique generated a 14% reduction in the average interpretation time, one third fewer interpretation errors, and two thirds fewer mouse clicks. The user satisfaction with our interface increased by 7% and 10% respectively, as reflected in our two usability questionnaires. Preliminary analysis also indicates eye-gaze over the controls was reduced in half.
8.1.7 DICOM WG11 and Supplement 60 on Hanging Protocols

Since 2001, through membership of DICOM WG11, we have been actively involved in the extension of the DICOM standard to provide support for image presentation for different types of viewing circumstances. In Section 4.9 we described our contribution to Supplement 60 to the DICOM Standard, which proposes solutions for storage, exchange, and query/retrieve of HPs. Our major contributions to Supplement 60 were in the area of image filtering and sorting, and in the consistent reproduction of colour images on different display devices.

8.2 Summary

We designed a task and a set of stimuli which allowed us to simulate the interpretation workflow from a typical radiological scenario. This was possible by abstracting the radiologist’s task and the basic workstation navigation functionality. We introduced an interaction technique attuned to the radiologist’s interpretation task, which generated significant performance gains, increased user satisfaction, and less eye-gaze over the controls.

The search for better interaction technique for the radiology workstation is just heating up. We think the silver bullet, if one exists, will come from a combination of software and hardware that will improve the interpretation process by significantly reducing off-image gaze fixations, and by optimizing the highly repetitive tasks involved in image navigation and manipulation. The steering wheel, the gas and the brake pedal\textsuperscript{16} are standard controls for cars in every price range, and accommodate drivers with all levels of experience. In this thesis we have started the search for a set of standard controls for the basic features of a radiology workstation that will accommodate all levels of expertise.

\textsuperscript{16} We mentioned using car pedals as alternate controls for the radiology workstation in an abstract submitted in 2002. We were surprised to hear on radiologist from UBC mentioning to us in April 2003 he would like to “use some pedals, like in the car” to interact with the radiology workstation.
8.3 Future work

Several suggested directions are listed that would continue this present research. Our analysis of the eye-tracking data from the pilot study was only used to assess the amount of eye gaze spent on the controls. We could extend our analysis to look for differences in the reading patterns for the two interaction techniques. We could also use the eye gaze information to classify the interpretation errors into search errors, pattern recognition errors, and decision making errors.

Eye-trackers may soon become reliable and simple enough that they will become yet another input device. By measuring the focus of both eyes, these devices will be able to pinpoint where one is looking in three-dimensional space. Interacting with computers via eye-movements could simplify one's immersion into the world of virtual reality.

The real estate problem may be alleviated with the use of 'focus plus context' screens, which are wall-size low-resolution displays with an embedded high-resolution display region. Focus plus context screens provide a single, non-distorted view, which is essential for the display of medical images. This technique could be especially useful to display dynamic information streams that are too large and detailed to fit on computer displays.

We are also interested in completing the external validation of these results with real radiologists, and thus showing the validation is robust across several interaction techniques. For example, we could compare subject’s performance using context-sensitive menus, invoked by a right mouse click, with traditional point-and-click on controls found on top of the screen which are becoming more cumbersome to use as monitors are getting bigger, thus requiring more mouse travel.

We could further explore the effect of differential learning and user fatigue. In order to assess the learning effects for each of the two interaction techniques presented, we could arrange for a follow up experiment with subjects performing two runs of the same interface. The effect of user fatigue could be tested by increasing the number of trials in each block from 15 to 30. Our subject’s performance could be investigated by checking how our subjects will tolerate the increase in workload. We hypothesize there
may not only be changes in the response time, but also in accuracy, user satisfaction, fatigue, eye movement patterns, and mouse travel.

We could also modify our controls for increased performance, targeting expert users who are more likely to use shortcut keys than to point-and-click on controls. It may be particularly interesting to assess the benefits of using FlowMenus, which allow fluid integration of consecutive menu selections, data entry, and direct manipulation. Novice users can perform a FlowMenu selection by popping up the radial menu and then selecting an item. More experienced users can make fast selections by drawing a straight mark in the direction of the desired menu item, eliminating menu popup. The advantages of FlowMenus are self-revelation, guidance and rehearsal. Self-revelation helps a novice determine what functions are available, while guidance and rehearsal train a novice to use the tools like an expert.

We are also interested in defining standard metrics for appraising radiology workstations. The workstation performance could be rated using a carefully selected set of representative use cases from radiological interpretation. These use cases should be taken into consideration based on the frequency of the radiological procedure, and the complexity of the interpretation as reflected by the Current Procedural Terminology. Several standardized workstation usability scores could be calculated using:

- a standard framework for analyzing routine human computer interactions, such as CPM_GOMS, which stands for Cognition, Perception, and Motor Goals, Operators, Methods and Selection Rules (John 1990). Assessing the disruptive effect of the workstation controls could be done automatically, thus generating absolute scores.

- an analysis of the disruptions of the visual search due to the manipulation of workstation controls, as shown in Section 4.8.1, and

- cognitive overload tests, such as the NASA Taskload Index. However, the scores on such tests are subjective, so they are only relevant as long as the same subjects were used during both experimental conditions.
Appendix A: First user study questionnaire

Filmless imaging vs. traditional light screen

1. How many computer image-viewers did you use and for how long?

2. Productivity: how many studies do you review per day. Each row signifies a typical day.

3. Quality of the diagnosis: which mode do you think is more accurate and why (use <, = and > between columns).

4. What fraction of the time spent on reviewing one study do you use for arranging the study?

Table 24: Productivity, Quality of diagnosis & Time spent arranging the study (%)

<table>
<thead>
<tr>
<th>Question</th>
<th>MR/CT</th>
<th>CR</th>
<th>RF</th>
<th>US</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of the diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time spent arranging the study (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. What are the major drawbacks of each mode? (enter in Table 25)
   - Example: (I) Small screen size; (II) lack of detail in context; (III) reduced resolution; (IV) reduced productivity; (V) personal preference.
6. What are the most used features? (enter in Table 25)
   o Example: (1) W/L; (2) Zoom; (3) Pan; (4) Magnifying Glass; (5) Cine play; (6) use previous studies for comparison; (7) linking series in stack mode.

7. What would be the most important features that you would want to have? (enter in Table 25)
   o Example: (a) Duplicate instances of a series with different W/L settings; (b) Hounsfield measurement on CT images; (c) enhanced printing capabilities; (d) enhanced mark & measure capabilities; (e) display healthy study example.

Table 25: Present Drawbacks, Most Used Features and Proposed New Features

<table>
<thead>
<tr>
<th>Modality</th>
<th>Drawbacks</th>
<th>Drawbacks</th>
<th>Most used features</th>
<th>New Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR/CT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Viewing protocols**

We propose implementing a viewing protocol like a collection of predefined settings (W/L, zoom, rotation, arrangement of the series in the viewports) that the system uses to display the images upon loading a study for review. It would be used to reduce the time spent to prepare a study for review. A viewing protocol template could be defined for each type of study, using a hierarchical architecture such as the below:
For plain film studies, such as CR, DR, and X-rays, you would most likely define protocols at the study level, where the same settings are used for all series in the study. These protocols contain general parameters such as the initial layout, viewing mode (stack or tile).

For cross-sectional studies, such as MR and CT, you may want to define protocols at the series level so that you can define the number of series within the study and then set the parameters, such as the window settings and orientation, for each series separately.

9. What other user preference should be stored in the viewing protocols and how important you consider this feature for each modality? Eg: size and color of annotations, W/L user-defined settings

<table>
<thead>
<tr>
<th>New feature vs. importance</th>
<th>All modalities</th>
<th>MR/CT</th>
<th>CR</th>
<th>RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. If templates for viewing preferences were to be defined (like ‘MRI of right knee without contrast’):
11. How many do you think you will need?
12. What percentage of studies do you think will fit in these templates (requiring minimal/no further adjustment)?

Table 27: Templates

<table>
<thead>
<tr>
<th></th>
<th>MR/CT</th>
<th>CR</th>
<th>RF</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Templates</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>
Displaying images

13. How important is the resolution (1728x2304 vs. 2048x2560)?
   - Size vs. resolution: show different images (from different modalities) at different layouts (7x9 MRI images fit in 1728x2304; they prefer a 2x2 layout, stack mode).

14. What is the preferred number of monitors for each resolution? Table (1/2/4 monitors vs. resolution)

15. Would you prefer a single big monitor instead of several smaller ones (‘frame effect’)?

16. How are these preferences dependent on the modality?
   - Find what is common and where/why they diverge: NM – color mapping; US – cine, color, sound; CT/MRI – grayscale Hi-res monitors.

17. Which modality you think best fits for each mode?

18. Can you fit all of the into a single Multi-Modality viewer?

Printing

19. DICOM

20. PS grayscale and color (annotations)

21. True size printing: studies only used true size X-ray images for diagnosis (potential subject for research).

22. Place patient demographic info on a banner on DICOM printers (format not supported; must pack all images into a single huge image and attach the header/footer to this one).
Appendix B: NASA Taskload Index (TLX)

Workload is a multi-dimensional psychological construct measuring the subjective experience of work that results from the mental actions performed while perceiving and processing information and executing a response.

The NASA TLX measures workload on 6 different dimensions (Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration) to create a picture of the amount and type of mental workload a user experiences during task performance (Staveland, Hart et al. 1986). It also combines subscale ratings that are weighted according to their subjective importance to the subject in a specific task rather than their a priori relevance to each subject's definitions of workload. This is determined by the subject's responses to pair-wise comparisons among the six factors (Hart 1987).

NASA TLX was developed by Sandra Hart et al at NASA Ames Research Center out of the earlier NASA Bipolar Rating Scale (Hart and Staveland 1988). The NASA TLX scale was created from 5 years of research on mental workload and task performance under a wide range of conditions conducted by many NASA and Academic researchers. It has been assessed in a variety of activities from simulated flight to simple laboratory experiments.

Other workload measurement questionnaires are:

- Subjective Workload Assessment Technique (SWAT),
- Modified Cooper Harper Scale (MCH), NASA RTLX,
- Subjective Mental Effort Questionnaire (SMEQ).

NASA TLX ratings have been shown to be very sensitive to experimentally manipulated levels of workload and substantially more reliable, as measured by a test-retest manipulation, than were SWAT ratings (Vidulich and Tsang 1986).

Byers et al (1988) compared NASA TLX to SWAT and MCH and found NASA TLX was both the most valid measure of subjective workload and had the highest user acceptance (Byers, Bittner et al. 1988). SWAT was second and MCH last. Also, NASA TLX was also found to correlate significantly with performance, and to correlate highly (0.75) with heart rate. The weighting technique succeeds in reducing between subjects variability more than any other commonly used subjective rating technique.

No useful reference values can be defined due to the fact that no reliable method for the definition of the experimental tasks is included. The use of reference values may be circumvented by applying the method within-subject to compare the subjective workload estimation between various situations.

**Description**

Although it is clear that definitions of workload do indeed vary among experimenters and among subjects (contributing to confusion in the workload literature and between-subject variability), it was found that the specific sources of loading imposed by different tasks are an even more important determinant of workload experiences. Thus, the current version of the scale (the Task Load Index) combines subscale ratings that are weighted according to their subjective importance to subjects in a specific task, rather than their a priori relevance to subjects' definitions of workload in general. NASA Task Load Index is a two-part evaluation procedure consisting of both weights and ratings.

**Sources of Load**

The first requirement is for each subject to evaluate the contribution of each factor (its weight) to the workload of a specific task. These weights account for two potential sources of between-subject variability: differences in workload definition.
between subjects within a task and differences in the sources of workload between tasks. In addition, the weights themselves provide diagnostic information about the nature of the workload imposed by the task. There are 15 possible pair-wise comparisons of the six scales. Subjects select the member of each pair that contributed more to the workload of that task. We tallied the number of times that each factor was selected. The tallies can range from 0 (not relevant) to 5 (more important than any other factor).

A different set of weights is obtained for each distinctly different task or task element upon its completion. The same set of weights can be used for many different versions of the same task if the contributions of the six factors to their workload are fairly similar. For example, the same set of weights was used for many different versions of a target acquisition task in which time pressure, target acquisition difficulty, and decision making load were varied. Obtaining separate weights for different experimental manipulations increased the sensitivity of the derived workload score only slightly, and did not warrant the additional time required to gather them. On the other hand, the weights obtained from the same subjects for a compensatory tracking task or a memory search task would not have been appropriate for the target acquisition task.

**Magnitude of Load**

The second requirement is to obtain numerical ratings for each scale that reflect the magnitude of that factor in a given task. The RATINGS sheet presents the six scales. Subjects use a pencil to mark each scale at the desired location. Each scale is presented as a line divided into 20 equal intervals anchored by bipolar descriptors (e.g., High/Low). Ratings may be obtained either during a task, after task segments, or following an entire task.

**Weighting and Averaging Procedure**

The overall workload score for each subject is computed by multiplying each rating by the weight given to that factor by that subject. The sum of the weighted ratings for each task is divided by 15, which is the sum of the weights. The user instructs the
program whether separate weights were collected for different subjects, experimental conditions, and replications.

**Subject Instructions**

Throughout this experiment the rating scales are used to assess your experiences in the different task conditions. Scales of this sort are extremely useful, but their utility suffers from the tendency people have to interpret them in individual ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended or the performance they achieved. Others feel that if they performed well, the workload must have been low and vice versa. Yet others feel that effort or feelings of frustration are the most important factors in workload and so on. The results of previous studies have found every conceivable pattern of values. In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

The evaluation you are about to perform is a technique developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort vs. Mental Demands) and asked to choose which of the items was more important to your experience of workload in the task(s) that you just performed. Each pair of scale titles will appear separately on the screen. Select the Scale Title that represents the more important contributor to workload for the specific task(s) in this experiment.
Table 28: TLX rating scale definitions

<table>
<thead>
<tr>
<th>Title</th>
<th>Endpoints</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>MENTAL DEMAND</td>
<td>Low/High</td>
<td>How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?</td>
</tr>
<tr>
<td>PHYSICAL DEMAND</td>
<td>Low/High</td>
<td>How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?</td>
</tr>
<tr>
<td>TEMPORAL DEMAND</td>
<td>Low/High</td>
<td>How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?</td>
</tr>
<tr>
<td>EFFORT</td>
<td>Low/High</td>
<td>How hard did you have to work (mentally and physically) to accomplish your level of performance?</td>
</tr>
<tr>
<td>PERFORMANCE</td>
<td>Good/Poor</td>
<td>How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?</td>
</tr>
<tr>
<td>FRUSTRATION LEVEL</td>
<td>Low/High</td>
<td>How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?</td>
</tr>
</tbody>
</table>
Appendix C: Examples on how various factors affect the trial complexity

Introduction: Shapes, targets, and distractors

Figure 68 and Figure 69 show all 8 shapes, all 4 potential targets, and respectively all 7 distractors used to synthesize our stimuli.

Target enclosure and target contrast (relative to background)

Figure 70 and Figure 71 show the effect of the presence of a frame around the shape: background has intensity 220, and the various shapes have intensities from 200-240.
Target location

The effect of the location of the potential\textsuperscript{17} target on image complexity is illustrated in Figure 72 and Figure 73.

Number of objects

The effect of the number of objects (potential targets and distractors) on image complexity is illustrated in Figure 74 and Figure 75.

\textsuperscript{17} Since the second image is not shown, one cannot verify the wild card.
Registration

The effect of the registration difficulty is illustrated below. Figure 76 and Figure 77 show the two images of a study, with difficult registration due to missing objects, and significant shift in object position. An example of a much easier registration is shown between Figure 78 and Figure 79.
Figure 78: Easy registration with the next image.

Figure 79: Easy registration with the previous image.
Appendix D: Training script for the user study

Training step 1: target definition

You are looking for a target in an image. A target is represented by two adjacent discs of the same diameter. Both discs are splinted in half vertically or both are splinted in half horizontally, with each half of different gray shades. Examples of targets and distractors are presented in Figure 1 and Figure 2, respectively.

Practice session 1

You are looking for a target in a study containing two related images. This time a splinted disc from the target previously described is replaced by a full disc that we will refer to as a "wild card."

A wild card can correspond to a target, or to a distractor.

When you look for a target in a study and you find a wild card on one image, you must look on the second image and find what this wildcard stands for.
A wild card is a disc with a uniform fill, and it holds the place for 
1. A disc with the proper cluster configuration, part of a target, as seen in Figure 1. 
2. A disc with another cluster configuration (distraction), as shown in Figure 2. 
3. A disc with one cluster, as it is a distraction, as shown in Figure 3.

Training step 3

Practice session 2 (contd.)

If there is a target in the following ‘study’ (a study is a pair of two similar images).

Practice session 2

Identify if there is a target in the following ‘study’ (a study is a pair of two similar images).

Practice session 2 (contd.)

Identify if there is a target in the following ‘study’ (a study is a pair of two similar images).

Important

A potential target will ALWAYS have a wild card. You must find on the second image what the wild card stands for, so you can differentiate a target from a distractor.

Designing Better User Interfaces for Radiology Workstations

182
Practice session 3
You are looking for a target in the following two (new and old) studies. You must:
1. report if there is a target in the new study;
2. report if there is a target in the old study;
3. in case a target is found on both studies (so it's the same target), report how the target's size evolved (remained the same, or grew, or shrunk).

Summary
The experiment consists of two sessions, and each session will have 15 trials. Each trial involves a search task for a target on two studies: an old study, and a new study.
- Each study has two images that represent similar views of the same scene. Each study has at most one target.
- If a target is found in both studies, it must be the same target (except for the target's size).
- For each individual trial, you must:
  1. report if there is a target in the new study;
  2. report if there is a target in the old study;
  3. in case a target is found on both studies (so it's the same target), report how the target's size evolved (remained the same, or grew, or shrunk).
Appendix E: System Usability Scale (SUS)

System Usability Scale is ‘quick-and-dirty’ usability scale, a 10-item questionnaire that gives an overview of satisfaction with software. It was developed by John Brooke, and is freely available for use providing acknowledgement is made of the source (Brooke 1992).

Similar usability questionnaires are:

- Satisfaction profile: SUMI: SUMI (Software Usability Measurement Inventory) is a 50 item questionnaire that measures five aspects of user satisfaction (Likability, Efficiency, Helpfulness, Control and Learnability), and scores them against expected industry norms. It can be purchased from HFRG.
- User interface satisfaction: QUIS (Questionnaire for User Interaction Satisfaction) is similar to SUMI, but measures attitude towards eleven interface factors (screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, on-line tutorials, multimedia, voice recognition, virtual environments, internet access, and software installation). It does not have industry norms. QUIS can be purchased from the University of Maryland.

SUS was developed in 1986 as part of the introduction of usability engineering to Digital Equipment Corporation’s integrated office systems program. Its objectives were to provide an easy test for subjects to complete (i.e. minimal number of questions), to be easy to score, and to allow cross-product comparisons. It has been used extensively in evaluations of projects in Digital Equipment Corporation (office systems, system management, technical tools and hardware systems) and has been found to be simple and reliable.
Using the SUS scale

The SUS scale is generally used after a user has had an opportunity to use a system but before any debriefing or discussion takes place. Users should be asked to record their immediate response to each item, rather than thinking about items for a long time. All items must be checked. If a user feels that they cannot respond to a particular item, they should mark the centre point of the scale.

Scoring the SUS scale

The SUS scale is a Likert scale and yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.

To calculate the SUS score, first sum the score contributions from each item. Each item’s score contribution will range from 0 to 4. For items 1, 3, 5, 7 and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU. SU scores have a range of 0 to 100.

Notes on construction of the SU scale

The SU scale was constructed from an original pool of 50 items. Ratings were made by 20 users on all 50 items for 2 systems, one designed for end-user use and one designed for use by systems programmers. These two systems were chosen to represent extremes of usability.

The 10 items selected were those which evoked the most consistent and most polarized responses. The items selected all have inter-correlations of between 0.7 and 0.9.
Table 29: System Usability Scale Questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th></th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think I would like to use this system frequently.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. I found the system unnecessarily complex.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. I thought the system was easy to use.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use this system.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. I found the various functions in this system were well integrated.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in this system.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to use this system very quickly.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. I found the system very cumbersome to use.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. I felt very confident using the system.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10. I need to learn a lot of things before I could get going with this system.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
## Appendix F: Second usability questionnaire (Q19)

<table>
<thead>
<tr>
<th>1. Overall, I am satisfied with how easy it is to use this system</th>
<th>strongly disagree</th>
<th>1 2 3 4 5 6 7</th>
<th>strongly agree</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. It was simple to use this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>3. I can effectively complete my work using this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>4. I am able to complete my work quickly using this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>5. I am able to efficiently complete my work using this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>6. I feel comfortable using this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>7. It was easy to learn to use this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>8. I believe I became productive quickly using this system</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>9. The system gives error messages that clearly tell me how to fix problems</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>10. Whenever I make a mistake using the system, I recover easily and quickly</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>11. The information (such as online help, on-screen messages, and other documentation) provided with this system is clear</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>12. It is easy to find the information I needed</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
<tr>
<td>13. The information provided for the system is easy to understand</td>
<td>strongly disagree</td>
<td>1 2 3 4 5 6 7</td>
<td>strongly agree</td>
<td>NA</td>
</tr>
</tbody>
</table>
14. The information is effective in helping me complete the tasks and scenarios  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |
15. The organization of information on the system screens is clear  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |
16. The interface of this system is pleasant  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |
17. I like using the interface of this system  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |
18. This system has all the functions and capabilities I expect it to have  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |
19. Overall, I am satisfied with this system  | strongly disagree | 1 2 3 4 5 6 7 | strongly agree | NA |

List the most negative aspect(s):

1. 
2. 
3. 

List the most positive aspect(s):

1. 
2. 
3. 

Designing Better User Interfaces for Radiology Workstations
Appendix G: Consent forms

First user study, November 1999

Consent for 'Presentation of Medical Images' Study

Adrian Moise

By signing either A or B below, I agree to participate in this ('Presentation of Medical Images') study. I understand that this means that I will answer questions that will provide feedback and my preferences with regard to various presentations of medical images on a computer screen. I further understand that my interaction with the workstation may be recorded on a camcorder.

If the material is published: (Choose A or B)

A) I wish to express my desire to be acknowledged. I understand that this means that my name will be mentioned in the article.

Date:__________________________________________________________

Signature:_____________________________________________________

Name in full:__________________________________________________

B) I wish to express my desire NOT to be acknowledged. I do not want my name associated in any way with the study.

Date:__________________________________________________________

Signature:_____________________________________________________

Name in full:__________________________________________________

Designing Better User Interfaces for Radiology Workstations
Pilot experiment

Consent for 'Image display for radiological softcopy interpretation' Study by Adrian Moise

By signing either A or B below, I agree to participate in this ('Image display for radiological softcopy interpretation') study. I understand that this means that I will answer questions that will provide feedback and my preferences with regard to various presentations of images on a computer screen. I further understand that my interaction with the workstation may be recorded using an eye-tracking device, and by having the entire session recorded on a camcorder.

I understand that I may withdraw my participation in this experiment at any time. I also understand that I may register any complaint I might have about the experiment with the researcher named above or with Dr. Ze-Nian Lee, Director of the School of Computing Science of Simon Fraser University. I may obtain copies of the results of this study, upon its completion, by contacting Adrian Moise at (604) 291 5509. I understand that my supervisor or employer may require me to obtain his or her permission prior to my participation in a study such as this.

If the material is published: (Choose A or B)

A) I wish to express my desire to be acknowledged. I understand that this means that my name will be mentioned in the article.

Date: ____________________________
Signature: _________________________
Name in full: _______________________

B) I wish to express my desire NOT to be acknowledged. I do not want my name associated in any way with the study.

Date: ____________________________
Signature: _________________________
Name in full: _______________________

Designing Better User Interfaces for Radiology Workstations
User study

SIMON FRASER UNIVERSITY

INFORMED CONSENT BY SUBJECTS TO PARTICIPATE

IN 'Image display for radiological softcopy interpretation' EXPERIMENT

The University and those conducting this project subscribe to the ethical conduct of research and to the protection at all times of the interests, comfort, and safety of subjects. This form and the information it contains are given to you for your own protection and full understanding of the procedures. Your signature on this form will signify that you have received a document which describes the procedures, possible risks, and benefits of this research project, that you have received an adequate opportunity to consider the information in the document, and that you voluntarily agree to participate in the project.

Having been asked by Adrian Moise of the Department of Computing Science of Simon Fraser University to participate in a research project experiment, I have read the procedures specified in this document. I understand this means that I agree to answer questions that will provide feedback and my preferences with regard to various presentations of images on a computer screen. I understand that my interaction with the workstation will be recorded on a camcorder. I understand that I may withdraw my participation in this experiment at any time.

I also understand that I may register any complaint I might have about the experiment with the researcher named above or with Dr. Ze-Nian Lee, Director of the School of Computing Science of Simon Fraser University. I may obtain copies of the results of this study, upon its completion, by contacting Adrian Moise at (604) 291 5509. I have been informed that the research material will be held confidential by the Principal Investigator. I understand that my supervisor or employer may require me to obtain his or her permission prior to my participation in a study such as this.

I agree to participate by signing this consent form.

Designing Better User Interfaces for Radiology Workstations
NAMES (please type or print legibly): ______________________________________

ADDRESS: __________________________________________________________________

__________________________________________________________

SIGNATURE: _____________________ WITNESS: _____________________

DATE: __________________________

ONCE SIGNED, A COPY OF THIS CONSENT FORM AND A SUBJECT FEEDBACK FORM SHOULD BE PROVIDED TO THE SUBJECT.
Appendix H: Approvals from the Office of Research Ethics

SIMON FRASER UNIVERSITY

OFFICE OF VICE-PRESIDENT, RESEARCH
BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6
Telephone: (604) 291-4370
FAX: (604) 291-4850

May 7, 1999

Mr. Adrian Moise
Graduate Student
School of Computing Science
Simon Fraser University

Dear Mr. Moise:

Re: Computer Presentation of Medical Images
NSERC

I am pleased to inform you that the above referenced Request for Ethical Approval of Research has been approved on behalf of the University Research Ethics Review Committee. This approval is in effect for twenty-four months from the above date. Any changes in the procedures affecting interaction with human subjects should be reported to the University Research Ethics Review Committee. Significant changes will require the submission of a revised Request for Ethical Approval of Research. This approval is in effect only while you are a registered SFU student.

In some cases, it is advisable that you obtain the permission of the employer of the subject involved prior to beginning your research study.

Best wishes for success in this research.

Sincerely,

Dr. Adam O. Horvath, Chair
University Research Ethics Review Committee

c: S. Atkins, Supervisor

/bjr

Designing Better User Interfaces for Radiology Workstations 193
October 10, 2002

Mr. Adrian Moise
Graduate Student
School of Computing Science
Simon Fraser University

Dear Mr. Moise:

Re: Image Display for Radiological Soft Copy Interpretation

I am pleased to inform you that the above referenced Request for Ethical Approval of Research has been approved on behalf of the Research Ethics Board. This approval is in effect for twenty-four months from the above date. Any changes in the procedures affecting interaction with human subjects should be reported to the Research Ethics Board. Significant changes will require the submission of a revised Request for Ethical Approval of Research. This approval is in effect only while you are a registered SFU student.

Your application has been categorized as 'minimal risk' and approved by the Director, Office of Research Ethics, on behalf of the Research Ethics Board in accordance with University policy R20.0, http://www.sfu.ca/policies/research/r20-01.htm. The Board reviews and may amend decisions made independently by the Director, Chair or Deputy Chair at its regular monthly meetings.

"Minimal risk" occurs when potential subjects can reasonably be expected to regard the probability and magnitude of possible harms incurred by participating in the research to be no greater than those encountered by the subject in those aspects of his or her everyday life that relate to the research.

Best wishes for success in this research.

Sincerely,

Dr. Hal Weinberg, Director
Office of Research Ethics

c: S. Atkins, Supervisor
References


Designing Better User Interfaces for Radiology Workstations 198
Designing Better User Interfaces for Radiology Workstations


Rogers, D. C., R. E. Johnston, et al. (1985). Predicting pacs console requirements from the radiologist's reading habits. SPIE.


*Designing Better User Interfaces for Radiology Workstations* 202


