MIXED REALITY INTERPERSONAL SIMULATION AFFORDS COGNITIVE, PSYCHOMOTOR, AND AFFECTIVE LEARNING

By

AARON ANDREW KOTRANZA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2009
To my wife, Sarah, for her love and support
ACKNOWLEDGMENTS

I thank my research advisor and supervisory committee chair, Dr. Benjamin Lok, for his advice, direction and support. I also thank my collaborators in medical education which made this work possible and provided valuable insight into real-world applications of this work: Dr. D. Scott Lind, Dr. Carla Pugh, Dr. Juan Cendan, Dr. Adeline Deladisma, and Andy Laserna. Thank you to Dr. Samsun Lampotang, Dr. Jorg Peters, Dr. Paul Fishwick, and Dr. Alireza Entezari for being my supervisory committee members, and for their ideas and support in my research. Kyle Johnsen, Andrew Raij, Brent Rossen, John Quarles, Xiyong Wang, Joon Chuah, and the rest of the past and present members of the Virtual Experiences Research Group provided invaluable assistance in conducting this research and were always there to discuss new ideas, no matter how off-the-wall. I thank the University of Florida alumni who provided financial support for my work through a University of Florida Graduate Alumni Fellowship. I thank my family for their love and support: my parents Steve and Rosemary, my brother Evan, my sister Alissa, and especially my ever encouraging wife, Sarah.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>12</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>14</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>17</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>18</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>22</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1 Motivation: Enhancing Educational Methods for Teaching Interpersonal Scenarios</td>
<td>25</td>
</tr>
<tr>
<td>1.1.1 Peer Simulation</td>
<td>26</td>
</tr>
<tr>
<td>1.1.2 Human Actor Simulation</td>
<td>28</td>
</tr>
<tr>
<td>1.1.3 Interpersonal Simulation with Virtual Humans</td>
<td>30</td>
</tr>
<tr>
<td>1.1.3.1 Current approaches to interpersonal simulation lack touch</td>
<td>31</td>
</tr>
<tr>
<td>1.1.3.2 Feedback of learner performance is limited</td>
<td>33</td>
</tr>
<tr>
<td>1.1.3.3 Learning and training transfer have not been demonstrated in current approaches to interpersonal simulation</td>
<td>34</td>
</tr>
<tr>
<td>1.2 Motivation: Augmenting Education of Medical Interpersonal Scenarios Underserved by Current Educational Approaches</td>
<td>35</td>
</tr>
<tr>
<td>1.2.1 Enhancing Intimate Exam Education</td>
<td>35</td>
</tr>
<tr>
<td>1.2.1.1 Intimate exams require touch for cognitive, psychomotor, and affective task components and require concurrent use of these three skill sets</td>
<td>36</td>
</tr>
<tr>
<td>1.2.1.2 Intimate exam training is underserved by existing educational approaches</td>
<td>37</td>
</tr>
<tr>
<td>1.2.1.3 Improving intimate exam training has potential for broad social benefit</td>
<td>41</td>
</tr>
<tr>
<td>1.2.2 Increasing Exposure to Abnormal Physical Findings</td>
<td>42</td>
</tr>
<tr>
<td>1.2.2.1 Training neurological examination with abnormal findings is underserved by existing educational approaches</td>
<td>42</td>
</tr>
<tr>
<td>1.2.2.2 Neurological examination requires touch and concurrent use of cognitive, psychomotor, and affective skill sets</td>
<td>44</td>
</tr>
<tr>
<td>1.2.2.3 Increasing exposure to abnormal findings in neurological exams has potential for broader social benefit</td>
<td>45</td>
</tr>
<tr>
<td>1.3 Thesis</td>
<td>46</td>
</tr>
<tr>
<td>1.4 Overview of Approach</td>
<td>46</td>
</tr>
<tr>
<td>1.4.1 Technological Innovation of MRIPS and Application to Medical Interpersonal Scenarios</td>
<td>47</td>
</tr>
</tbody>
</table>
1.4.2 Evaluation of the Validity of MRIPS for Practicing and Evaluating Cognitive, Psychomotor, and Affective Task Components ........................................48
1.4.3 Incorporation or Real-time Feedback of Cognitive, Psychomotor, and Affective Task Performance ........................................................................48
1.4.4 Evaluation of Learning and Training Transfer ........................................50
1.5 Innovations .....................................................................................................53

2 REVIEW OF LITERATURE ..............................................................................58

2.1 Foundations for Interpersonal Simulation with Virtual Humans ...................58
2.1.1 Social Responses to Virtual Humans .........................................................59
2.1.2 Toward Changing Human Behavior ........................................................60

2.2 Interpersonal Simulation with Virtual Humans .............................................62
2.2.1 Current Approaches to Interpersonal Simulation .......................................62
2.2.2 Existing Interpersonal Simulators ..............................................................66
2.2.3 Incorporation of Feedback in Interpersonal Simulation ..............................68

2.3 Motivation for Touch in Interpersonal Simulation .......................................69
2.3.1 The Role of Touch in Communication .......................................................70
2.3.2 Touch in Virtual Environments .................................................................71

2.4 Other Approaches to Medical Interpersonal Simulation .............................73

3 MRIPS DESIGN PRINCIPLES AND DEVELOPMENT OF MRIPS-CBE .............75

3.1 MRIPS Design Principles ...........................................................................76
3.2 Clinical Breast Examination ........................................................................79
3.2.1 CBE Procedure .......................................................................................79
3.2.2 Cognitive, Psychomotor, and Affective Components .................................82
3.2.3 Current Approaches to Teaching Clinical Breast Examination .................83

3.3 MRIPS-CBE ..................................................................................................85
3.3.1 Motivations and Goals ................................................................................85
3.3.2 Merging and Augmenting in MRIPS-CBE ...............................................86
3.3.3 Cognitive, Psychomotor, and Affective Affordances of MRIPS-CBE ..........87

3.4 Visual Interface ............................................................................................88
3.4.1 Life-sized Virtual Human Characters .......................................................89
3.4.2 Augmenting the Virtual World with Real Objects .....................................89
3.4.3 Display of the Visual Interface ................................................................92
3.4.3.1 Head-Mounted Display (HMD) ...........................................................93
3.4.3.2 Projection display ................................................................................93
3.4.4 Perspective Correct Viewing of the Visual Interface .................................94
3.4.5 Registering Visual and Physical Interfaces ...............................................94

3.5 Physical Interface ........................................................................................95
3.5.1 Active Sensing of User Touch ..................................................................95
3.5.2 Passive Detection of User Touch and Manipulation of Tools and Props ......97
3.5.3 Bidirectional Touch: Enabling the Virtual Human to Touch the User ........99
3.5.3.1 Purely virtual touch ............................................................................99
3.5.3.2 Physical touch ......................................................................................100
5.4 The Haptic Interface ........................................................................................................ 162
  5.4.1 Prior Approaches ........................................................................................................ 162
  5.4.2 Haptic Interface: Wii-Remote and External Sensing ............................................. 163
  5.4.3 Virtual Hand-Held Tools and Hand Gestures ............................................................ 165
    5.4.3.1 Ophthalmoscope .................................................................................................. 165
    5.4.3.2 Eye chart .............................................................................................................. 166
    5.4.3.3 Hand gesture tool ................................................................................................ 167
  5.4.3 The Haptic Interface Enhances Communication in Interpersonal Simulation ............ 170

5.5 Usability and Content Validity of MRIPS-NEURO for Practicing Diagnoses of Abnormal Findings ............................................................................................................ 171
  5.5.1 Study Design and Procedure ...................................................................................... 171
  5.5.2 Results ........................................................................................................................ 172
  5.5.3 Observations ............................................................................................................... 174
  5.5.4 Conclusions and Continued Evaluation ...................................................................... 175

6 REAL-TIME EVALUATION AND FEEDBACK OF PERFORMANCE ............. 184
  6.1 Motivation for Feedback .................................................................................................. 185
  6.2 Unique Capabilities of MRIPS to Evaluate Performance and Provide Feedback ............ 186
  6.3 Choice of the Visual Channel to Provide Feedback ....................................................... 189

7 FEEDBACK IN MRIPS-CBE ......................................................................................... 191
  7.1 Introduction ...................................................................................................................... 191
    7.1.1 Cognitive Components ............................................................................................ 191
    7.1.2 Affective Components ............................................................................................. 192
    7.1.3 Psychomotor Components ....................................................................................... 193
  7.2 Procedural Checklist ....................................................................................................... 194
  7.3 Thought Bubbles ............................................................................................................. 195
    7.3.1 Automated Evaluation of Affective Performance .................................................... 197
    7.3.2 Feedback to Reinforce and Correct Affective Performance ....................................... 199
  7.4 Touch Map ...................................................................................................................... 200
    7.4.1 Feedback Goals ....................................................................................................... 200
    7.4.2 Capturing Palpation Pressure and Pressure in an Expert’s CBE ............................... 201
    7.4.3 Guiding and Evaluating Complete Coverage .......................................................... 203
    7.4.4 Calculating the Palpation Pressure Levels ................................................................ 204
    7.4.5 Design of Feedback Elements to Guide, Reinforce, and Correct ............................... 208
    7.4.6 Presenting Feedback In-situ with the Virtual Human and Physical Breast ................... 210
    7.4.7 Design Choices ........................................................................................................ 210
      7.4.7.1 How many experts are needed to model psychomotor performance? .......................... 210
      7.4.7.2 Visual feedback elements occlude the learner’s hands ........................................ 211
      7.4.7.3 Drawbacks of an image-based approach ............................................................. 211
  7.5 Pattern-of-Search Map .................................................................................................. 212
7.5.1 Feedback Goals ........................................................................................ 212
7.5.2 Modeling Correct Pattern-of-Search .......................................................... 212
7.5.3 Guiding and Evaluating Learner Pattern-of-Search .................................... 213
7.5.4 Design of the Feedback Elements for Guiding, Reinforcement, and Correction ........................................................................................................ 215
7.6 Post-Experiential Feedback ............................................................................... 216
7.7 Face Validity of Touch Map and Pattern-of-Search Map Feedback ................. 217

8 FEEDBACK IN MRIPS-NEURO ........................................................................ 230

8.1 Introduction ......................................................................................................... 230
8.2 H-Map ................................................................................................................. 230
8.3 Patient Vision ...................................................................................................... 232
  8.3.1 Feedback Goals ........................................................................................ 232
  8.3.2 Prior Work in Motivating Perspective Taking ........................................... 233
  8.3.3 Patient Vision Feedback ........................................................................... 234
8.4 Evaluating the Impact of Feedback on Cognitive, Psychomotor, and Affective Performance ........................................................................................... 235
  8.4.1 Study Design and Procedure ................................................................... 235
  8.4.2 Population ................................................................................................. 238
  8.4.3 Metrics ....................................................................................................... 238
    8.4.3.1 Evaluating cognitive and affective performance ............................ 238
    8.4.3.2 Evaluating psychomotor performance ............................................ 239
  8.4.3 Hypotheses ............................................................................................... 241
  8.4.4 Results and Discussion ............................................................................ 242
    8.4.4.1 Hypothesis Patient-Vision improves affective. Experiencing patient vision increases concern for patient safety: accepted .......................... 242
    8.4.4.2 Hypothesis Patient-Vision improves cognitive. Experiencing patient vision aids diagnosis of CN disorder: rejected but with a positive result ................................................................. 244
    8.4.4.3 Hypothesis H-Map improves psychomotor completeness. H-Map visualization results in a more complete eye movements test: rejected ......................................................................................... 245
    8.4.4.4 Hypothesis H-Map improves psychomotor efficiency. H-Map visualization results in a more efficient eye movements test: accepted 249
  8.4.5 Conclusions ............................................................................................... 250

9 LEARNING, TRAINING TRANSFER, AND IMPACT OF REAL-TIME FEEDBACK IN MRIPS-CBE ........................................................................ 258

9.1 Introduction ......................................................................................................... 259
9.2 Study Design ...................................................................................................... 260
  9.2.1 Evaluating Learning and Training Transfer .................................................. 260
  9.2.2 Evaluating the Impact of Real-Time Feedback on Performance .................. 262
  9.2.3 Control Groups for Investigating the Validity of Study Results .................. 263
9.3 Population .......................................................................................................... 264
9.4 Statistical Analysis ............................................................................................. 265
9.5 Cognitive Performance ........................................................................................................ 266
9.5.1 Measures ....................................................................................................................... 267
  9.5.1.1 Breast history completeness .................................................................................. 267
  9.5.1.2 Visual inspection completeness ......................................................................... 269
9.5.2 Hypotheses .................................................................................................................... 269
9.5.3 Results: Breast History Learning and Training Transfer ............................................. 270
9.5.4 Results: Impact of Feedback on Breast History Completeness ............................... 274
9.5.5 Visual Inspection Learning and Training Transfer ....................................................... 276
9.5.6 Discussion .................................................................................................................... 278
9.6 Psychomotor and Cognitive-Psychomotor Performance .................................................. 278
9.6.1 Measures ....................................................................................................................... 279
  9.6.1.1 Coverage and correct pressure ........................................................................... 279
  9.6.1.2 Correct pattern of search .................................................................................... 282
  9.6.1.3 Finding masses ..................................................................................................... 283
9.6.2 Hypotheses .................................................................................................................... 284
9.6.3 Results: Coverage and Pressure Learning and Transfer ............................................. 285
  9.6.3.1 Coverage learning .............................................................................................. 285
  9.6.3.2 Coverage transfer ............................................................................................... 287
  9.6.3.3 Pressure learning ............................................................................................... 288
  9.6.3.4 Pressure transfer ............................................................................................... 289
9.6.4 Results: Impact of Real-Time Feedback on Coverage and Pressure .......................... 290
9.6.5 Results: Pattern-of-Search Learning and Transfer ...................................................... 293
9.6.6 Results: Impact of Real-Time Feedback on Pattern-of-Search ................................. 297
9.6.7 Results: Finding Masses and False Positives Learning and Transfer ....................... 298
9.6.8 Discussion .................................................................................................................... 301
9.7 Affective Performance ..................................................................................................... 302
9.7.1 Measures ....................................................................................................................... 302
9.7.2 Hypotheses .................................................................................................................... 306
9.7.3 Results: Empathy Learning ....................................................................................... 306
9.7.4 Results: Impact of Feedback ..................................................................................... 307
9.7.5 Results: Empathy Transfer ....................................................................................... 309
9.7.6 Discussion .................................................................................................................... 310
9.8 Validity of Results .......................................................................................................... 311
  9.8.1 Impact of Multiple MRIPS Practice Opportunities .................................................... 311
  9.8.2 Impact of an SP Pre-test Interaction on Subsequent SP Performance ..................... 312
9.9 Study Limitations .......................................................................................................... 313
9.10 Revisiting Meta-Hypotheses ......................................................................................... 314

10 SUMMARY AND FUTURE DIRECTIONS ........................................................................... 330
  10.1 Review of Results ......................................................................................................... 330
  10.2 Future Directions ......................................................................................................... 331

APPENDIX

A STUDY MRIPS-SP DATA .................................................................................................. 334
A.1 Breast History Checklist Data ................................................................. 334
A.2 Empathy Video Review Instrument ....................................................... 334
A.3 Empathy Video Review Data ................................................................. 335

B STUDY MRIPS-x2 DATA ............................................................................ 337
B.1 Study MRIPSx2 Video Review Instrument ............................................ 337
B.2 Study MRIPSx2 Video Review Data ..................................................... 337
B.3 Study MRIPSx2 Breast History Checklist Data ...................................... 338
B.4 Study MRIPSx2 Palpation Completeness Data ....................................... 338

C STUDY MRIPS-NEURO QUESTIONNAIRES ........................................ 339
C.1 Study MRIPS-NEURO Post-Patient Vision Survey ................................. 339
C.2 Study MRIPS-NEURO Post-Exam Survey ............................................. 339

D STUDY MRIPS-LEARNING INSTRUMENTS AND DATA .................... 341
D.1 Study MRIPS-Learning Breast History Checklist Data ........................ 341
D.2 Study MRIPS-Learning Coverage and Pressure Data .......................... 342
D.3 Study MRIPS-Learning Pattern-of-Search Data .................................... 343
D.4 Study MRIPS-Learning Empathy Video Review Instrument ............... 343
D.5 Study MRIPS-Learning Empathy Video Review Data .......................... 344
D.6 Affective Ratings of Participants in MRIPS-CBE ................................. 346
D.7 Pilot Study Video Rating Instrument and Data .................................. 346

LIST OF REFERENCES ................................................................................. 349

BIOGRAPHICAL SKETCH .......................................................... 363
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Affordances and drawbacks of prior educational approaches and MRIPS for teaching cognitive, psychomotor, and affective aspects of high-stakes interpersonal scenarios</td>
<td>55</td>
</tr>
<tr>
<td>1-2</td>
<td>Traditional and simulation approaches to teaching intimate exams</td>
<td>56</td>
</tr>
<tr>
<td>1-3</td>
<td>Feedback in MRIPS</td>
<td>56</td>
</tr>
<tr>
<td>3-1</td>
<td>List of simulation inputs and outputs</td>
<td>114</td>
</tr>
<tr>
<td>3-2</td>
<td>Using system state to direct the conversation between user and MRH</td>
<td>114</td>
</tr>
<tr>
<td>4-1</td>
<td>Items in the medical history completeness checklist</td>
<td>148</td>
</tr>
<tr>
<td>4-2</td>
<td>Results of video review of critical moments</td>
<td>148</td>
</tr>
<tr>
<td>4-3</td>
<td>Population breakdown for Study MRIPSx2</td>
<td>149</td>
</tr>
<tr>
<td>4-4</td>
<td>Cognitive performance results for Study MRIPSx2</td>
<td>149</td>
</tr>
<tr>
<td>4-5</td>
<td>Psychomotor performance results for Study MRIPSx2</td>
<td>149</td>
</tr>
<tr>
<td>5-1</td>
<td>List of cranial nerves which can be examined using MRIPS-NEURO</td>
<td>176</td>
</tr>
<tr>
<td>5-2</td>
<td>Focused neurological exam tasks</td>
<td>176</td>
</tr>
<tr>
<td>5-3</td>
<td>Usability ratings of MRIPS-NEURO</td>
<td>176</td>
</tr>
<tr>
<td>9-1</td>
<td>Instrument used to evaluate the completeness of breast history taking in MRIPS and SP interactions</td>
<td>318</td>
</tr>
<tr>
<td>9-3</td>
<td>Performance in breast history taking in Study MRIPS-Learning</td>
<td>319</td>
</tr>
<tr>
<td>9-4</td>
<td>Changes in the number of participants asking about specific risk factors</td>
<td>319</td>
</tr>
<tr>
<td>9-5</td>
<td>Number of participants performing any visual inspection and complete visual inspections in each interaction</td>
<td>319</td>
</tr>
<tr>
<td>9-6</td>
<td>Instrument used to evaluate coverage and use of correct pressure in the SP interactions</td>
<td>320</td>
</tr>
<tr>
<td>9-7</td>
<td>Summary of acceptance and rejection of hypotheses of psychomotor and cognitive-psychomotor task performance</td>
<td>320</td>
</tr>
</tbody>
</table>
Coverage and use of deep pressure of the Study MRIPS-Learning participants .......................................................... 321

Total deviation from expert pattern-of-search in the three MRIPS interactions of Study MRIPS-Learning. ......................................................... 321

Number of participants finding real masses and false positive masses in MRIPS. ........................................................................................................ 321

Participants finding masses palpated a larger percentage of the MRIPS breast with deep pressure ............................................................ 322

Affective performance in MRIPS-CBE interactions .......................................................................................................................... 322

Expert ratings of participants’ affective performance in the SP interactions ...... 322

Performance in pilot study used to assess impact of a single MRIPS-CBE interaction on a subsequent CBE of an SP .................................................. 323

Performance in a CBE of an SP after three MRIPS-CBE practice opportunities and after one MRIPS-CBE practice opportunity .......................... 323

Concurrent improvement in the three skill sets ................................................. 324
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>The simulator of Pugh et al. [38]</td>
<td>57</td>
</tr>
<tr>
<td>3-1</td>
<td>A learner performs a CBE in MRIPS-CBE</td>
<td>115</td>
</tr>
<tr>
<td>3-2</td>
<td>System design of MRIPS-CBE</td>
<td>116</td>
</tr>
<tr>
<td>3-3</td>
<td>MRIPS-CBE visual interface presented on a projection screen.</td>
<td>117</td>
</tr>
<tr>
<td>3-4</td>
<td>The physical interface of MRIPS-CBE</td>
<td>117</td>
</tr>
<tr>
<td>3-5</td>
<td>The first approach taken to tracking the physical gown and providing a corresponding virtual gown used a background subtraction approach</td>
<td>118</td>
</tr>
<tr>
<td>3-6</td>
<td>Affording bidirectional touch for communication by allowing the MRH to touch the user</td>
<td>119</td>
</tr>
<tr>
<td>3-7</td>
<td>The physical interface incorporates a mechanical right arm, allowing active-haptic touch from virtual human to human user</td>
<td>120</td>
</tr>
<tr>
<td>3-8</td>
<td>The layers of the framework for abstracting a virtual environment's control of physical actuators</td>
<td>120</td>
</tr>
<tr>
<td>3-9</td>
<td>Progression of one time step of the simulation module</td>
<td>121</td>
</tr>
<tr>
<td>3-10</td>
<td>The medical interview portion of the interaction with the MRH breast exam patient</td>
<td>121</td>
</tr>
<tr>
<td>3-11</td>
<td>Two of the poses required for visual inspection</td>
<td>122</td>
</tr>
<tr>
<td>3-12</td>
<td>The xml script that defines relationships between servos, constraints, and “animation” of servos to allow the MRH to touch the user on the hand.</td>
<td>122</td>
</tr>
<tr>
<td>4-1</td>
<td>Appearance of the MRIPS-CBE patient for Study MRIPS-SP</td>
<td>150</td>
</tr>
<tr>
<td>4-2</td>
<td>The appearance of the MRIPS-CBE mixed reality human in Study MRIPSx2</td>
<td>151</td>
</tr>
<tr>
<td>4-3</td>
<td>Visualization of a participant’s CBE completeness</td>
<td>151</td>
</tr>
<tr>
<td>5-1</td>
<td>An expert performs a neurological exam of Vic, a virtual human patient with double vision due to CN6 palsy</td>
<td>177</td>
</tr>
<tr>
<td>5-2</td>
<td>The cardinal eye movements of a normal, unaffected eye</td>
<td>177</td>
</tr>
<tr>
<td>5-3</td>
<td>Cardinal movements with the left eye affected by CN3 palsy</td>
<td>178</td>
</tr>
</tbody>
</table>
5-4 Cardinal movements with the left eye affected by CN6 palsy. ..............................179

5-5 The cardinal vectors for normal, CN3, and CN6 eye movements are graphed as (yaw, pitch) pair associated with each axis .................................................................180

5-6 A) Testing the pupillary reflex with the ophthalmoscope.  B) Performing the fundoscopic test with the ophthalmoscope. ........................................................................181

5-7 Visual acuity test with the virtual eye chart. ................................................................181

5-8 The finger counting test .......................................................................................... 182

5-9 Checking the eye movement of a virtual human patient with a left eye affected by CN6................................................................. 182

5-10 Testing facial sensitivity by touching the virtual human’s face. ..............................183

7-1 The breast history portion of the procedural checklist is displayed above the virtual human’s head ...............................................................................................220

7-2 The visual inspection portion of the procedural checklist expands to show the three poses required for visual inspection .................................................................221

7-3 The procedural checklist also incorporates feedback to aid in the cognitive task of recalling which peripheral areas of lymph nodes should be examined 221

7-4 Thought bubble feedback when the learner responds with empathy .....................222

7-5 Thought bubble feedback when the learner responds inappropriately ...................223

7-6 The touch map provides feedback of coverage and use of correct palpation pressure .................................................................................................................224

7-7 The pairing of the color and infrared seeing cameras and the haptic interface to the virtual human .................................................................................................224

7-8 A) The boundary of the area required for complete coverage of the breast cone. B) Complete coverage is indicated when this area is filled. .................................225

7-9 Informal correctness of the model is demonstrated by showing that the output of the model fits the expected progression of pressure levels .................................225

7-10 The color of the feedback provides guidance, reinforcement, and correction of the learner’s palpation pressure through real-time changes in color .................226

7-11 A learner follows an expert’s vertical strip pattern-of-search .....................................226

7-12 Modeling pattern-of-search ..................................................................................227
The touch map and pattern-of-search map for the same exam

Feedback is provided concerning cognitive elements such as the procedure of visual inspection and cognitive-psychomotor elements such as completeness of palpation.

Summary feedback of affective performance.

Progression of the H-Map visualization as the learner performs the eye movement test.

The VSP feedback experience of Raij et al.

The initial implementation of patient vision alpha blended each eye's image to present double vision on a non-stereoscopic display.

Study procedure.

Physical setup of the study.

Views during the exam.

Procedure for Study MRIPS-Learning.

Participants' performance in breast history completeness in the two SP and three MRIPS interactions.

Real-time feedback appears to be more effective than the post-experiential feedback.

A) Participant drawing of correct locations of the two masses in the MRIPS breast. B) Participant drawing of correct location of the mass in the SP breast.

Participants' use of deep pressure in MRIPS.

Patterns-of-search closely following and significantly deviating from an expert.
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBE</td>
<td>Clinical Breast Examination. Refers both to the exam itself and the act of performing the exam. A Clinical breast examination involves conversing with the patient to take a breast history, visually inspecting the patient’s breasts, and palpating the patient’s breasts to search for abnormalities.</td>
</tr>
<tr>
<td>MRIPS</td>
<td>Mixed Reality Interpersonal Simulation. Our expansion of interpersonal simulation to incorporate touching of the virtual human, manipulation of hand-held tools, and real-time feedback of performance.</td>
</tr>
<tr>
<td>CN</td>
<td>Cranial nerve. Refers to one of the twelve cranial nerves.</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MIXED REALITY INTERPERSONAL SIMULATION AFFORDS COGNITIVE,
PSYCHOMOTOR, AND AFFECTIVE LEARNING

By
Aaron Andrew Kotranza

December 2009

Chair: Benjamin Lok
Major: Computer Engineering

In fields such as medicine, military, and law enforcement, success in interpersonal
scenarios requires mastering a complex group of cognitive, psychomotor, and affective
skills. These interpersonal scenarios involve tasks which call upon multiple skill sets
concurrently. For example, in a clinical breast exam, the doctor must concurrently recall
a pattern of palpation (cognitive) while palpating with correct pressure (psychomotor)
and also keeping a patient comforted (affective). The concurrent use of these skill sets
mandates that learners practice concurrent actuation of these skills; each skill set
should not be learned in isolation.

However, traditional education approaches provide a level of practice that is
inadequate for many novices to achieve competency in the three skill sets and to learn
how to concurrently call on multiple skill sets to succeed in the interpersonal scenario.
This is due to reasons such as lack of standardization and inability to recreate these
interpersonal scenarios with peers or human actors.

Interpersonal simulation with virtual humans promises on-demand, highly
immersive learning experiences which could augment these curricula to provide
learners the additional practice required to achieve competency. For this reason,
interpersonal simulation is receiving increased attention from early adopters in medicine, military, and law enforcement fields.

However, interpersonal simulation has yet to be deployed as part of curricula in these fields. We have identified two shortcomings of current approaches to interpersonal simulation: the absence of touch – which contributes to training of all three skill sets, and a lack of evidence that learning and skills transfer takes place in users of the simulations.

To address the shortcomings of current approaches, we introduce a new approach to interpersonal simulation – mixed reality interpersonal simulation (MRIPS). MRIPS incorporates passive and active haptic interfaces instrumented with sensors to afford touch between human and virtual human as well as the manipulation of hand-held tools. Thus, in addition to bidirectional verbal and gestural capabilities of traditional interpersonal simulation, MRIPS provides touch from human to virtual human, touch from virtual human to human, and touch from hand-held tools to virtual human. These touch inputs and outputs are used to affect the psychomotor, cognitive, and affective components of the simulated interpersonal scenario.

The incorporation of these haptic interaction capabilities in MRIPS addresses problems with prior approaches to interpersonal simulation, expanding the applicability of interpersonal simulation to training cognitive, psychomotor, and affective skills and their concurrent use:

1. **Problem:** Psychomotor task components involving touch could not be simulated.
   a. **Approach:** MRIPS incorporates touch from human to virtual human and from hand-held tool to virtual human to afford the simulation of psychomotor task components involving touch.
2. **Problem:** Prior interpersonal simulation approaches do not afford touch for communication between the human and virtual human. Touch is an essential component of communication between two humans and the lack of touch results in incomplete communication in these interpersonal simulators.
   a. **Approach:** By incorporating touch from human to virtual human and from virtual human to human, MRIPS affords interpersonal communication to accomplish both cognitive task components, e.g. achieving compliance in getting a patient to assume a specific pose for a physical exam, and affective components, e.g. a comforting touch.

3. **Problem:** Prior approaches to interpersonal simulation afforded only simple vision-based gesture interfaces, e.g. pointing and iconic gestures, or encumbering gesture interfaces, e.g. body suits.
   a. **Approach:** MRIPS incorporates instrumentation, e.g. six degree-of-freedom tracking, of the haptic interfaces. This affords simulation of hand-held tools for performing complex psychomotor task components not involving touch. These tools also serve as grounding objects to enhance communication between the human and virtual human.

4. **Problem:** Feedback and reflection motivated by feedback is necessary for learning. Prior interpersonal simulations could not provide feedback of psychomotor performance without encumbering interfaces, e.g. gloves, body suits or expert observers.
   a. **Approach:** MRIPS incorporates instrumentation of the non-encumbering haptic interfaces for touching the virtual human and for tool manipulation. This affords quantitative measuring of the learner’s psychomotor, cognitive, and affective skills performance as well as real-time feedback to guide and elicit reflection on the learner’s performance. By providing real-time feedback of learner performance, MRIPS increases the potential for learning in the simulated interpersonal scenario.

We applied MRIPS to simulate two interpersonal scenarios in medicine – the clinical breast exam and the neurological exam – which could not be simulated through prior interpersonal simulation approaches. User studies established the validity of MRIPS for practicing and evaluating learners' psychomotor, cognitive, and affective skills. We then incorporated visual feedback of user performance in these three skill sets, to enhance the potential for learning. Additional user studies were then conducted to determine what learning occurs in users of MRIPS and whether skills learned in MRIPS transfer to the real-world scenarios being simulated.
Results show that MRIPS affords psychomotor, cognitive, and affective learning required for success in real-world interpersonal scenarios and that skills learned in MRIPS translate to the real world scenarios. This work demonstrates the validity of using MRIPS to train real-world interpersonal scenarios and motivates further incorporation of MRIPS into interpersonal skills curricula.
CHAPTER 1
INTRODUCTION

Success in interpersonal scenarios depends on one’s ability to draw from a complex mix of cognitive, psychomotor, and affective skills. By acquiring these skill sets and mastering the concurrent actuation of these skills, learners attain competency in these scenarios.

However, there is increasing evidence that the traditional curricula tasked with teaching these interpersonal scenarios fail, for a significant\(^1\) percentage of students, in bringing students to competency [1][2][3]. Coupled with inadequate opportunities for evaluating students, this results in graduates of these curricula lacking skills necessary to succeed in the interpersonal scenarios being taught [4][5].

Instead, competency in these scenarios is achieved through “trial-by-fire” in later apprenticeship stages of education, e.g. the residency following medical school. It is the goal of educators to bring learners to competency before the learners are placed in these real-world interpersonal scenarios in which failure carries with it high risks, e.g. of harming a patient, malpractice lawsuits. The shortcomings of traditional curricula in achieving this goal are due to the limited set of training tools available to educators. Lectures, practice with peers and actors, and purely-physical simulation (e.g. anatomical models) all have qualities that limit their availability and effectiveness at training concurrent use of cognitive, psychomotor, and affective skills [4][6][7][8][9].

\(^1\) For the example scenario of the clinical breast exam, the percentage of students graduating without achieving competency varies by experiment and scenario component: 83% of 4th-year students reported needing additional training in CBE [2]; 65% of 1st-year residents failed a standardized patient assessment of CBE performance [3]; graduating students failed to use a correct pattern-of-search 55% of the time, failed to perform visual inspection 25% of the time, and found only 40% of masses present [1].
My work seeks to provide educators with an additional tool, one that is autonomous, always available, and able to concurrently train all three skill sets. This tool is mixed reality interpersonal simulation (MRIPS), a novel approach to simulation of interpersonal scenarios that merges haptic interfaces, physical sensing, and virtual humans, and is effective in training the cognitive, psychomotor, and affective components of interpersonal scenarios.

MRIPS provides passive and active haptic interfaces to a life-sized virtual human agent. The haptic interfaces are instrumented with sensors which provide quantitative measures of manipulation of the haptic interfaces. This allows the virtual human to respond to the manipulating of these haptic interfaces and affords automated evaluation of learners’ cognitive, psychomotor, and affective performance.

This is a fundamentally new approach to interpersonal simulation with virtual humans – the simulation of an interpersonal scenario by replacing one’s interaction partner with a virtual human. Prior approaches to interpersonal simulation have allowed users to communicate with the virtual human using speech (bidirectional verbal communication) and limited gestures (e.g. pointing) and for the virtual human to communicate using facial expressions and gestures [10][11]. MRIPS provides a richer, more complete (as compared to human-human interaction) set of interaction capabilities. In addition to bidirectional verbal communication and unidirectional gestural communication from the virtual human, MRIPS provides:

- Bidirectional touch from human to virtual human and from virtual human to human. Touch is enabled for cognitive and affective aspects of communication as well as for psychomotor task performance.
- Hand-held tool manipulation including touching of the virtual human using the hand-held tools.
• Robustly recognized user hand-gestures to communicate with the virtual human and serve as conversational grounding elements.

• Real-time feedback of the learner’s performance in cognitive, psychomotor, and affective components of the scenario, including feedback to guide affective aspects of communication, such as expressions of empathy.

These affordances of MRIPS uniquely enable concurrent learning of the cognitive, psychomotor, and affective components of an interpersonal scenario – and skills learned in MRIPS transfer to the real-world interpersonal scenario. This dissertation presents the development and evaluation of MRIPS in the following progression:

1. Design of haptic interfaces to enable touch and hand-held tool use in the interpersonal simulation, as a means of performing the cognitive, psychomotor, and affective components of interpersonal scenarios. Application of these interfaces to two interpersonal scenarios in the medical domain: a clinical breast exam and a neurological exam with abnormal findings.

2. Evaluation of MRIPS to establish the validity of MRIPS for practicing the cognitive, psychomotor, and affective components of these medical examination scenarios. This evaluation focuses on demonstrating that learners treat the examination of a virtual human using MRIPS similarly to how they treat the examination of a human patient. Learners’ performances in MRIPS are shown to be similar in quality to their exams of human patients. Additionally, learners of different levels of experience with the real-world interpersonal scenario are able to perform in MRIPS in a manner consistent with their experience level. These evaluations establish the validity of MRIPS as an additional practice opportunity to augment underserved aspects of traditional curricula. Establishing this validity sets the groundwork for evaluating learning within MRIPS and training transfer of skills learned in MRIPS to real-world interpersonal scenarios.

3. Enhancement of the educational capabilities of MRIPS through the addition of novel real-time and post-experiential feedback of quantitatively measured learner performance. Feedback has been shown to be necessary for learning in the scenarios to which MRIPS is targeted [12]. The novel feedback capabilities of MRIPS are designed to elicit reflection and result in improvement of cognitive, psychomotor, and affective task components. We present results of user studies which demonstrate that this feedback improves learners’ performance in these three skill sets.

4. Evaluation of the efficacy of MRIPS for cognitive, psychomotor, and affective learning, and the training transfer of learned skills to the real-world interpersonal scenario.
1.1 Motivation: Enhancing Educational Methods for Teaching Interpersonal Scenarios

Interpersonal simulation is the recreation of a real-world interpersonal scenario by substituting a virtual human for one’s interaction partner. Interpersonal simulation has been proposed to train high-stakes interpersonal scenarios with the goal of bringing novices to a level of competency that will allow them to succeed in the real-world scenario. Examples of high-stakes scenarios are medical examination and military negotiation. These scenarios are characterized by the quality that failure is both unacceptable and highly likely without having previously achieved competency [10][13]. Interpersonal simulation is well suited for practicing these high-stakes scenarios because it presents a learning environment in which it is acceptable to fail [14]. This quality allows learners to potentially undergo the progression of failure, feedback, reflection, and improvement which is critical to attaining competency [12], before being exposed to the real-world scenario. As a result, practicing with interpersonal simulation is expected to provide learners with an increased chance of success in the high-stakes real-world scenario [13][15].

The driving force behind using interpersonal simulation to practice these scenarios is the finding that learning these high-stakes scenarios is underserved by traditional educational methods [4][12]. Traditional approaches to interpersonal scenario education involve lecture or book-based learning, followed by a “learning by doing” experience. This experience is a re-creation of the interpersonal scenario with a peer or human actor [16]. Interpersonal simulation seeks to recreate these experiences with a virtual human in place of the peer or human actor. The affordances and drawbacks of each approach are listed in Table 1-1.
1.1.1 Peer Simulation

Simulating an interpersonal scenario using a peer as the interaction partner affords practice of communication, including the cognitive components of progressing through a set of questions to ask or topics to converse about and responding correctly to the peer, psychomotor components such as touch for communication and physical examination, and affective components such as perspective taking and expressing empathy.

However, using peers to practice psychomotor components involving touch is limited to a set of scenarios for which touch of another is ethical and does not make either participant uncomfortable. While psychomotor tasks such as suturing a simulated wound on a suturing pad worn by the peer (as used in [7]) can be practiced, touching of the peer is not acceptable in intimate examinations, especially in mixed gender peer pairs. Surveys of medical students have found that a majority of students of both genders are uncomfortable with having intimate exams performed on them by a peer [17][18]. For these high-stakes interpersonal scenarios, practice with peers has largely been replaced by practice with human actors [6]. Even for non-intimate exams, practice with peers was rated as the fourth most effective approach to learning physical examination in a survey of 83 American medical schools. Peer learning was surpassed by exams of standardized human patients (human actors) and real patients, as well as observing expert exams of human actors [6].

Another drawback of practicing with peers is that the experience is often (~33% of the time)\(^2\) not treated seriously, especially in embarrassing and awkward medical scenarios.

\(^2\) There is no widely reported measure of how seriously role-playing with peers is taken. From a limited sample of published studies, role-playing with peers is a negative experience for roughly one-third of
physical exam scenarios. This causes the affective components of the scenario to not be effectively practiced [8]. Other shortcomings of this approach are that abnormal findings and cultural differences can not be simulated. This limits the set of scenarios to which the peer approach can be applied [19].

Abnormal physical findings, specific to the medical domain, are physical abnormalities or pathologies present in the interaction partner, e.g. a lazy eye. Practicing interpersonal scenarios with a partner of a different culture is a goal of cultural competency training. Examples include addressing cultural differences between European-American and African-American patients [20] and dealing with non-native English speakers [21] or criminal suspects with a mental handicap [22]. Simulation of abnormal findings and cultural differences with peers is difficult simply because peers with the desired abnormalities and peers from diverse cultural backgrounds are rarely available. With respect to cultural differences, this is partially due to the demographics of medical students. From 2006 to 2008, more than 64% of matriculants of United States medical schools were Caucasian, with Asians (both East and South Asian) the next largest group at just under 20% [23]. When culturally diverse peers are available, the use of these peers in physical examination scenarios is further limited due to cultural and religious beliefs [19]. The drawback of peer simulation for teaching cultural competencies is readily apparent in situations in which the medical student (peer) population has a different cultural makeup than the surrounding patient population. For example, at our collaborating institution, the Medical College of

students. Commonly reported reasons for the negative experience are a lack of realism and inability or unwillingness to become engaged (i.e. take the experience seriously). One study reported that role-play with peers is the teaching method least preferred by 32% of novice medical students [24]. Another study in support of role play reported that 22% of participants had prior negative role-play experiences; however, this number actually represented 33% of participants who had prior role-play experiences [8].
Georgia, the majority of the student population is Caucasian, while the majority of the population of the surrounding city and county is African-American [25]. The military domain is one in which peers are actively recruited for cultural competency training [26], e.g. the recruitment of native Arabic speakers as peers for training scenarios including negotiating with civilians. Seeking out, compensating, and training an individual to serve as an interaction partner describes the “human actor” approach.

1.1.2 Human Actor Simulation

In the human actor approach, an interaction partner is sought who possesses a specific set of physical traits. This actor is then trained to portray an individual of a specific group. For example, medical educators might recruit a post-menopausal female to portray a patient with multiple risk factors for breast cancer, e.g. family history, smoking, hormone replacement use. In medical education, human actors are known as standardized human patients (SPs). SPs are the gold-standard for providing medical students with practice in medical interview and examination [16].

This approach has many of the same affordances and drawbacks as the peer approach, but provides a more standardized approach. The actor is more standardized than the peer, as the same actor may be used for an entire group of learners, while the peer group typically divides a group pair-wise. The actor is also trained from a script developed by educators. The recruiting of actors with specific backgrounds or traits allows cultural differences to be simulated. The experience is taken seriously by learners and has been validated in many domains as an equivalent substitution for the real-world scenario [16]. The actor is compensated monetarily for a specific scenario, so issues related to touching and psychomotor tasks are eliminated, e.g. a female SP may have a breast exam conducted by medical students regardless of student gender.
Drawbacks of the human actor approach are largely logistical: finding an actor with specific characteristics, maintaining availability of the actor for an entire group of learners, and compensating the actor [6].

Because of the issue of availability, the inability to simulate abnormal findings persists with the human actor approach. It is rare to find a person with the desired abnormality, such as a breast mass, who is willing to be subjected to examination by a class of medical students. Some abnormalities can not be simulated with actors because the abnormalities are life threatening or require immediate treatment. An example of this category of abnormal finding is double vision due to a cranial nerve abnormality. Additionally, it is logistically difficult to find actors for certain groups such as the elderly or minorities. It may be unethical to recruit other specific patient types such as children or people with psychological disorders. Some institutions have policies preventing the use of actors for intimate exam scenarios because of privacy concerns [6].

These logistic issues also extend to issues of standardization. Actors are costly and each actor may be available for a limited set of hours, resulting in a limited number of practice opportunities for learners. The same actor is unlikely to be available for an entire group of learners (each U.S. medical school had an average of ~135 1st-year students enroll in 2008 [23]). An actor’s performance may vary depending on his or her mood, and some actors will follow the script more closely and act more convincingly than other actors. For these reasons, a group of learners may receive highly varying experiences, making evaluation of the learners’ performances difficult (e.g. in a medical interview scenario, medical educators would have to normalize learner scores based on
how easily an SP divulged her medical history). Howley provides a review of these issues and prior evaluations of the efficacy of SPs. Even with these drawbacks, human actors are widely used in training medical interpersonal scenarios. Seventy-five percent of medical schools use SPs to teach or evaluate examination skills [27].

1.1.3 Interpersonal Simulation with Virtual Humans

Interpersonal simulation with virtual humans was developed to address the drawbacks of the human actor approach. As the virtual human interaction partner is an autonomous agent, the simulation is always available. The virtual human can be programmed to follow a script more closely than a human actor, providing a greater degree of standardization and less variance in the learners’ experiences [14]. Virtual humans can also be developed to present abnormal findings such as double vision, a breast mass, or a facial burn, and can be given specific cultural traits for cultural competency training [20][28].

For these reasons, interpersonal simulation has received attention from early adopters in fields training high stakes interpersonal scenarios, including medical interview, informed consent, and examination [15][29][30], mental health [31][32], military leadership and negotiation [13][21][33], and law enforcement training [22]. However, interpersonal simulation has yet to be incorporated into curricula to train real end-users in these fields. We have identified three main shortcomings of current approaches to interpersonal simulation which may contribute to the lack of curriculum integration: 1) current approaches to interpersonal simulation universally lack touch between human and virtual human, 2) feedback of learner performance is not widely used and is primarily limited to post-experiential review (after action review), and 3)
there is a lack of evidence that learning and skills transfer takes place in users of existing interpersonal simulations.

1.1.3.1 Current approaches to interpersonal simulation lack touch

Touch is a crucial part of the cognitive, psychomotor, and affective components of interpersonal scenarios. Touch contributes to both communication and the performance of psychomotor tasks.

Within communication, touch is used in both affective and cognitive components. For example, touch can be used to comfort or express an emotional connection (affective) or to instruct the interaction partner’s movement or to achieve her compliance with one’s instructions (cognitive) [34][35][36]. Without providing interfaces for touch between human and virtual human, current interpersonal simulations can not fully simulate communication between two humans. MRIPS provides bidirectional touch – touch from human to virtual human and from virtual human to human. The addition of touching affords a set of communication modalities more similar to real-world interpersonal scenarios.

In addition to communicating by touching of one’s communication partner, touch contributes to communication by enabling manipulation of other objects in the environment. Manipulated objects serve as a common ground for communication between interaction partners. A common ground is a pool of mutually agreed upon information which serves as a way to ensure that a verbal or nonverbal message intended to be communicated is received by the other communication partner [37]. For example, in a neurological examination scenario, an ophthalmoscope manipulated by the doctor serves as a grounding object. The doctor asks the patient to “follow the light with your eyes.” Because the patient sees that the doctor is holding an
ophthalmoscope, the patient knows that “the light” refers to the light of the ophthalmoscope, and the patient follows the ophthalmoscope as it is moved by the doctor. Grounding is especially important in communication between human and virtual human because of the error present in verbal interfaces such as speech recognition. Even with hours of speech recognition training, recognition of a user’s speech is imperfect. In the same neurological exam scenario, if the doctor asks “follow the light with your eyes” the speech recognition may produce text such as “follow the lead with your tires” (actual output from Microsoft speech recognition SDK 5.1). Without the presence of grounding objects, there is not enough information present in this nonsensical utterance to allow the virtual human to respond appropriately. However, if the virtual human knows that the doctor is holding an ophthalmoscope with the light turned on, the keyword “follow” provides enough information for the virtual human to know to follow the position of the ophthalmoscope with his eyes.

In addition to contributing to communication, touch is used to perform tasks with psychomotor components. These tasks may be purely psychomotor, for example, palpating (touching) a clinical breast exam patient’s breast with specific movements of the hand. Other tasks in which touch is used may require concurrent cognitive and psychomotor skills. Examples of this class of task include recalling (cognitive) a cultural greeting with elements of touch and speech and enacting (psychomotor) this greeting when meeting a foreign business client, or palpating (psychomotor) a breast with a hidden breast mass while recognizing (cognitive) whether the tissue feels like fatty tissue, fibrous tissue, or an abnormality. We refer to these as compound tasks. These compound tasks require the use of touch to perform a psychomotor task concurrently.
with a cognitive and/or affective task. An example of a compound psychomotor-affective task is comforting a nervous patient during a physical examination. Compound tasks may draw on all three skills sets, for example suturing an injured soldier while keeping him calm while instructing the soldier in how to apply pressure to the wound.

Novices typically perform poorly at compound tasks because they have not attained competency in all three skill sets and have not had adequate practice actuating all three skill sets concurrently [38]. Current approaches to interpersonal simulation can not provide the necessary practice to achieve competency in these joint tasks, as they lack interfaces that afford touch between human and virtual human. Thus the set of real-world interpersonal scenarios to which current interpersonal simulations can be applied is limited by the lack of touch.

A main innovation of this dissertation is the design of haptic interfaces to life-sized virtual humans which enable touching of the virtual human. These interfaces expand on the capabilities of prior approaches to interpersonal simulation, affording touch for enhancing communication and performing psychomotor and compound tasks. The design of these haptic interfaces is explored in the implementation of the breast exam and neurological exam simulations in Chapters 3 and 5.

1.1.3.2 Feedback of learner performance is limited

There are few instances of feedback of learner performance in existing interpersonal simulations. Existing approaches at providing feedback are typically limited to post-experiential feedback, e.g. the after-action review system of Raij et al. [84]. The existing approaches at providing feedback of performance in interpersonal simulation are discussed in Section 2.2.3. Prior work investigating real-world interpersonal scenarios such as medical physical examination has found that learners
prefer real-time feedback to post-experiential feedback. Additionally, feedback should be coupled to learner actions, i.e. learner actions should elicit changes in the feedback in real-time [12]. MRIPS expands on existing approaches to providing feedback of learner performance in interpersonal simulation by: 1) providing both real-time and post-experiential feedback; 2) providing feedback that guides, reinforces, and corrects performance in cognitive, psychomotor, and affective performance; and 3) coupling real-time feedback to learner actions, e.g. in the clinical breast exam, palpating the patient’s breast results in visual changes in the area of the breast palpated, to indicate correctness of the pressure used in palpation.

1.1.3.3 Learning and training transfer have not been demonstrated in current approaches to interpersonal simulation.

In addition to lacking touch, the efficacy of current approaches to interpersonal simulation to train the cognitive, psychomotor, and affective components of real-world interpersonal scenarios has not been evaluated. The goal of interpersonal simulation is the transfer of learned skills to real-world interpersonal scenarios – yet this has not been demonstrated.

In the literature, there are instances in which interpersonal simulation has elicited short-term changes in user behavior (discussed in Section 2.1.2). However, I was not able to find any published instances of learning the skills required for a real-world interpersonal scenario using interpersonal simulation, or improvement in a real-world interpersonal scenario due to practice with interpersonal simulation.

A main innovation of this dissertation is to determine whether learning occurs as a result of practice with mixed reality interpersonal simulation. We specifically evaluate whether a novice learner’s cognitive, psychomotor, and affective skills improve with
repeated use of MRIPS, and evaluate whether skills learned in MRIPS transfer to the real-world interpersonal scenario. This is explored in Chapter 9.

**1.2 Motivation: Augmenting Education of Medical Interpersonal Scenarios Underserved by Current Educational Approaches**

We are motivated to choose applications for MRIPS which: 1) require touch for cognitive, psychomotor, affective, and compound task components; 2) require concurrent use of skills from these three skill sets; 3) are underserved by existing educational approaches; and 4) may provide broader societal benefit as a result of enhancing educational approaches.

Along with collaborators at the Medical College of Georgia and University of Florida College of Medicine, we have evaluated two areas of medical education that can benefit from MRIPS: intimate physical examination and physical examination with abnormal findings.

Although both applications are medical physical exams, each has distinct sets of cognitive, psychomotor, and affective components and distinct uses of touch.

**1.2.1 Enhancing Intimate Exam Education**

Intimate exams are physical exams of intimate areas of the patient, e.g. female breast exams, male and female pelvic exams, and digital rectal exams. Due to their intimate nature these exams are anxiety provoking for both the patient and healthcare provider [38]. Because of this, affective components such as understanding the patient’s emotional state and empathizing with the patient are especially important. The physical examination itself is also especially difficult to learn and perform in the context of the stressful, anxiety-provoking patient encounter. The combination of an anxious novice learner with the difficult exam (e.g. 76% of experienced clinicians feel they need
to improve in breast examination [4]) causes novices to be unable to communicate effectively while simultaneously performing incomplete and incorrect exams [38].

We have focused specifically on clinical breast examination (CBE), as practice and evaluation of CBE skills can benefit from the unique combination of haptics and sensing provided by MRIPS. A critical component of CBE is palpating the breast at three distinct levels of pressure. The correctness of this palpation pressure can only be determined with sensing [39]. In our application of MRIPS to simulating CBE, MRIPS-CBE, MRIPS is able to provide realistic feeling simulated breast tissue and sensing of correct examination skills including correct palpation pressure. A unique aspect of MRIPS is to provide these affordances within the broader context of the affective and cognitive tasks of communicating with a patient.

1.2.1.1 Intimate exams require touch for cognitive, psychomotor, and affective task components and require concurrent use of these three skill sets

All intimate exams rely on touch extensively for communication and psychomotor task performance. As intimate exams are anxiety provoking for the patient (and the practitioner), touch is commonly used to comfort the patient [35]. Touch is also relied upon for attaining patient compliance and for instructing the patient, e.g. in clinical breast examination, to assume specific positions for visual inspection and palpation [34][36]. Specific to the clinical breast exam, scenario components can be classified as cognitive, psychomotor, affective, and compound cognitive-psychomotor. These components typically take place concurrently.

- **Cognitive**: 1) recalling of a series of questions to assess the patient’s risk of breast cancer and 2) thinking of appropriate verbal and nonverbal responses to patient statements and questions.
- **Psychomotor:** palpating the patient’s breast in circular motions at three levels of pressure, superficial (low), medium, and deep (high), without palpating at a “too-hard” level of pressure that may cause the patient pain.

- **Affective:** 1) keeping track of the patient’s verbal and nonverbal cues as to her emotional state, and 2) in response to patient emotions, alleviating the patient’s anxiety, addressing her concerns, and expressing empathy when appropriate. In order to foresee when empathy is appropriate, learners must gain some understanding of the patient’s perspective by engaging in perspective taking [40][41].

- **Cognitive-psychomotor:** Recalling of a procedural pattern, the pattern-of-search, in which the breast should be palpated and maintaining this pattern of palpations. While progressing through a series of palpations, recognizing which areas of the breast remain to be palpated to ensure palpation of the entire breast. At each palpation, interpreting whether the breast tissue feels like normal tissue or an abnormality.

- **Other compound tasks:** These include affective-psychomotor and cognitive-psychomotor tasks such as keeping the patient comforted during anxiety-provoking moments such as visual inspection of the patient’s breasts, opening the patient’s gown, and palpating the breast.

1.2.1.2 Intimate exam training is underserved by existing educational approaches

Educators note that clinical breast examination is difficult to learn and teach due to novice learner’s anxieties, the complex set of task components, and the difficulty in providing feedback of learners’ performance [9][38]. Typical approaches to practicing other physical exams, such as practice with human actors (standardized human patients), can not be provided frequently enough and do not provide the precise feedback (e.g. of palpation pressure, pattern of search) needed to achieve competency. SPs provide high-level objective ratings of cognitive and psychomotor performance (e.g. “did the student palpate the cone of the breast, yes or no”, “did the student perform a complete review of systems, yes or no”) and subjective rating of affective performance

---

3 At many institutions, including those we have partnered with to evaluate MRIPS, standardized patients are used primarily to evaluate learners, typically once at the end of the first year and then again before graduation.
(rating “the student encouraged the patient to express emotions” using a Likert scale). Some SPs are trained to provide more precise feedback, such as a qualitative, subjective determination of whether correct deep pressure was used in palpation. Regardless of the precision, there is no widely used standardized feedback [9].

The approach most widely accepted as effective for teaching CBE is to have an expert to observe the novice perform the exam on a human patient and for the expert to provide feedback as to the quality of the novice’s exam. However, this expert observation can not be provided frequently due to the high demand on the expert’s time, and some students graduate without being evaluated in performing a CBE. The current approaches to teaching CBE and intimate exam training in general allow medical students to graduate without a rigorous assessment of their competency in intimate exams. Without having received adequate practice opportunities and feedback, these graduating medical students express low confidence in their intimate exam skills [4][9].

Simulation approaches have addressed the shortcomings of traditional educational methods for intimate examination, but have their own drawbacks, which are listed alongside traditional approaches in Table 1-2.

Purely physical simulation, i.e. silicone anatomical models, provides learners with on-demand practice of physical exam skills and exposure to abnormal findings (e.g. simulated breast masses incorporated into the model). However, the on-demand nature of this approach does not extend to providing feedback on the quality of learner performance. Feedback is typically provided only when used in peer-to-peer teaching sessions incorporated into course curricula. The feedback received from a peer may not be the same quality available from an expert. Notably, purely physical simulation
does not provide the patient context (conversing with a patient) required for practicing cognitive and affective interpersonal skills.

Standardized human patients (SPs) provide high-level feedback (e.g. for CBE, the SP may notify the learner that he has missed the lymph nodes near the breast). Other than real patients, SPs provide the highest fidelity simulation of interpersonal skills. SPs have been validated as a replacement for real patient interactions in this regard [16]. However, SPs are rarely able to present abnormal findings, e.g. a breast mass or a lazy eye.

To provide exposure to abnormal findings, hybrid approaches have SPs wear the silicone anatomical models, such that the models appear to be part of the SP’s body. This provides exposure to abnormal findings within the patient context needed for interpersonal skills practice [7]. However, this approach retains the drawbacks of SPs related to availability (Table 1-1).

Pugh et al. have augmented silicone anatomical models with physical sensors to create simulators for female pelvic, rectal, and breast exams. This simulation approach provides students with immediate, objective feedback of their physical exam performance. However, this approach lacks the patient context needed for interpersonal skills practice. As with all approaches which do not afford practice of the cognitive and affective components involved in these interpersonal skills, this approach does not provide an opportunity to practice the concurrent actuation of cognitive, psychomotor, affective skill sets. Feedback in this “physical sim + sensing” approach takes the form of a series of meters and charts indicating correctness and completeness of exam technique (Figure 1-1). This presentation was designed to provide experts with
a more detailed (i.e. quantitative) method of evaluating learner performance. However, this feedback is not presented in a form meaningful to novice learners.

Approaches which do not simulate the interpersonal aspects of intimate exams are often used in conjunction with SP interactions. Typically the physical simulator is used to first train the cognitive and psychomotor components of physical examination. This is followed by SP interactions to train the other cognitive, psychomotor, and affective components of the interpersonal scenario. This separation hampers the learner’s ability to draw from all three skill sets concurrently, which is required for success in the real-world scenario.

MRIPS addresses the drawbacks of each approach. Similar to hybrid simulation, MRIPS provides opportunities to practice concurrent use of cognitive, psychomotor, and affective skill sets, but expands on this approach with the addition of on-demand availability and objective, quantitative, more precise feedback of student performance. Expanding on Pugh’s approach of incorporating sensing with physical simulation, MRIPS also simulates interpersonal components of the scenario, and provides more detailed feedback in a form expected to be more meaningful to novice learners (e.g. providing color-coded feedback at palpation at the three levels of pressure, vs. providing a meter from 0 volts to 5 volts).

A main innovation of this dissertation is the creation of automated, real-time feedback and evaluation of a learner’s cognitive, psychomotor, and affective performance. Surveys of novice medical students have shown that 70% of students learn best from experiences that combine visual and kinesthetic information. Expert observation favors auditory learners (a small minority of the population) as these
students can be “talked through” the exam [42]. MRIPS provides visual feedback in combination with kinesthetic information from the haptic interfaces, potentially providing a tool that can educate a large segment of medical (and perhaps other) students underserved by existing traditional and simulation approaches. Of the approaches detailed in Table 1-2, MRIPS is the only approach which simulates both the interpersonal and physical exam components of the intimate exam while also targeting visual-kinesthetic learners.

1.2.1.3 Improving intimate exam training has potential for broad social benefit

Clinical breast exam and other intimate exams such as the prostate exam are essential components of screening for early detection of several cancers and competence of intimate exams is a critical skill for all healthcare professionals [4][9]. CBE may find up to 10% of cancers that are not detectable with imaging techniques such as mammography [44]. Breast cancer is the most common form of cancer in women and is the second most common cause of cancer death [45]. Without CBE, approximately 10,000 breast cancers might otherwise go undetected each year [4]. If MRIPS-CBE is shown to be successful in training CBE and is incorporated into medical curricula and remediation programs, MRIPS-CBE has the potential to enhance early detection of cancers. This can potentially improve the quality of life for thousands of patients. Improving cognitive and affective skills through additional practice with MRIPS has the potential to improve patient outcomes through making better testing and treatment decisions and forging closer (i.e. friendlier, less businesslike) doctor-patient relationships.
1.2.2 Increasing Exposure to Abnormal Physical Findings

The second application area chosen for MRIPS is simulation of abnormal physical findings that cannot be simulated using traditional or existing simulation approaches. One such scenario is a neurological exam of a patient with cranial nerve palsy. Patients with cranial nerve palsy present with abnormal physical findings such as: a pupil that does not contract, an eye that does not move through a full range of motion, or double vision [46].

1.2.2.1 Training neurological examination with abnormal findings is underserved by existing educational approaches

The diagnosis of the cranial nerve palsy is based primarily on interpreting these abnormal findings in the context of the patient’s medical history [47]. Currently, medical students learn diagnosis through lecture, textbook, video-based instruction, and supervised patient encounters [48]. Exposure to abnormal findings in human patients is not standardized and occurs only if a student happens to be in the neurological clinic at the time that a patient arrives with abnormal findings, i.e. exposure is “catch as catch can.” Thus medical students may graduate without experiencing abnormal findings in a neurological exam. Students’ diagnostic skills are typically tested by an expert observer at most once in a neurology clerkship, and some students go untested. When these skills are tested, students arrive at (what an expert would consider) a correct diagnosis 50% of the time [49].

The lack of exposure to abnormal findings may also hamper affective skills with patients presenting with these abnormal conditions. Patients are often fearful because of social or cosmetic problems (e.g. lazy eye) and because they worry that they will lose sight completely [50] and their symptoms may indicate a serious condition such as
aneurysm or brain tumor [51]. Without the experience of talking to patients with these fears and taking these patients’ perspective, novice learners may not address the patient’s concerns or consider the patient’s safety (e.g. did a patient with severe double vision drive to the clinic?). It is imperative for the physician to address issues of patient safety [52]. As with intimate exams, the limited opportunities for practice and evaluation causes students to report a low level of knowledge of the neurological exam and low confidence in their abilities [53][54].

Simulation approaches to providing additional practice of the neurological examination include purely physical and purely virtual simulation.

Purely physical simulation has yet to completely recreate abnormal findings in a neurological examination. Recreation of abnormal findings such as restricted eye movements and partial loss of sensation in the face would require robotics beyond that used in sophisticated physical simulators such as the Human Patient Simulator [55]. Instead of attempting to simulate a full exam, physical simulation has focused on training narrower components such as diagnosing abnormal findings in the fundoscopic exam (looking at the patient’s retina through the ophthalmoscope). Slides of photos of abnormal retinas are inserted in the eyes of a mannequin head [56].

Purely virtual simulation approaches have more completely simulated the neurological exam and abnormal findings. A purely virtual web-based approach (2D graphics) has been developed by the University of California Davis [57] to train diagnosis of cranial nerve palsies. A pair of disembodied eyes follows the cursor, providing the user with information to make a diagnosis based on eye movements.
There are currently no simulation approaches which address a complete set of abnormal findings (i.e. eye internal and external appearance, movement, and reaction to light; drooping eyelid, sensation or lack of sensation in the face and eyes; expression of seeing double based on where the eyes are looking) or which simulates the interpersonal aspects of the scenario. Our application of MRIPS to simulating a neurological exam with abnormal findings, MRIPS-NEURO, represents the first attempt to recreate this interpersonal scenario in a manner that affords practice of the cognitive, psychomotor, and affective skills required by the scenario. Users of MRIPS-NEURO are currently able to perform tests required to diagnose disorders of seven cranial nerves: 2, 3, 4, 5, 6, 7, and 12 (there are a total of 12 cranial nerves).

1.2.2.2 Neurological examination requires touch and concurrent use of cognitive, psychomotor, and affective skill sets

A neurological exam with abnormal findings requires touch (including hand-held tool use) for cognitive, psychomotor, and affective components. These components occur concurrently, requiring concurrent actuation of the three skill sets. The exam components of a neurological exam are many [54][58]. Those which are required for diagnosing cranial nerve palsies are:

- **Cognitive:** Conducting a medical history. Recalling various tests and interpreting the results of the tests: testing pupillary reflex by shining the light of the ophthalmoscope into the patient’s eyes, examining the fundus (retina) using the ophthalmoscope, checking for double vision by holding fingers up and asking the patient how many fingers he sees, have the patient read from an eye chart to test visual acuity, shake a finger in the patient’s peripheral vision to test for peripheral vision disorders, ask the patient to blink, wink, stick out his tongue, turn his head from side to side, and touch his chin to his chest.

- **Psychomotor:** Testing the patient’s eye movements by moving a finger or light in the shape of an uppercase “H” with a size and shape that tests the extremes of the patient’s eye movements: far left, far right, far upper-left, lower-left, upper-right, and lower-right. Touching the patient’s face with a finger to test for loss of sensitivity in the face.
• Affective: Alleviating the patient’s anxiety and addressing the patient’s concerns, through engaging in taking the patient’s perspective and expressing empathy when appropriate. Additionally the user must consider issues related to patient safety [52].

Although all of the tests performed in the neurological exam have psychomotor components, from discussion with medical experts, two tasks are considered to be more difficult. These are the fundoscopic exam, which requires skilled use of the ophthalmoscope, and the eye movements test. Notably, MRIPS-NEURO does not have the goal of training ophthalmoscope use – practice with a real ophthalmoscope and peer is a higher fidelity and more commonly available method for learning how to use the ophthalmoscope. MRIPS-NEURO focuses less on training psychomotor components than MRIPS-CBE, instead focusing primarily on the integration of information gained from the many simple tests and medical history to arrive at a correct diagnosis.

The one psychomotor task which MRIPS-NEURO seeks to train is learning the “H” pattern to test eye movements. The size and shape of the “H” vary with the depth of the user’s finger to the patient’s eyes – with experience, clinicians learn to perform this exam up close and have memorized the movements that result in the patient’s eyes moving to the extremes (the six endpoints of the “H”). MRIPS-NEURO may be able to assist novice learners in memorizing these movements in addition to providing novices with increased exposure to patients with abnormal findings.

1.2.2.3 Increasing exposure to abnormal findings in neurological exams has potential for broader social benefit

The majority of patients presenting with neurological disorders are not first seen by a neurologist, but instead by a general practitioner or emergency room clinician [59]. Up to twenty percent of acute medical admissions are neurological disorders – these
patients require immediate treatment. Thus, all healthcare practitioners need to be able to recognize the symptoms of a variety of cranial nerve disorders [60]. By increasing medical students' exposure to and practice diagnosing abnormal findings in a neurological exam, MRIPS-NEURO may improve patient outcomes in a scenario that is underserved by current educational approaches.

An innovation of this dissertation is the application of MRIPS to simulating clinical breast examination and neurological examination with abnormal findings, providing an additional tool in the educator’s arsenal for educating future generations of medical students and maintaining the skills of current residents and clinicians.

1.3 Thesis

The thesis deals with the novel integration of haptic interfaces and physical sensing with virtual human simulation and real-time feedback to, as a whole, initiate learning of the cognitive, psychomotor, and affective skill sets of an interpersonal scenario and demonstrate improvement in the real-world interpersonal scenario being simulated.

Thesis statement: Interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance improves users’ scenario-specific psychomotor, cognitive, and affective skills. Skills improvement transfers to the real-world interpersonal scenarios being simulated, demonstrated as improved performance in the real-world interpersonal scenario.

1.4 Overview of Approach

To investigate the truth of this thesis statement, we focused on innovating technology and evaluating its impact on learners, and finally evaluating learning and training transfer:

1. **Technological innovation:** Two interpersonal simulators incorporating instrumented haptic interfaces were designed: MRIPS-CBE and MRIPS-NEURO.
**Evaluation:** Two user studies (total of 52 participants) were conducted to demonstrate that learners’ cognitive, psychomotor, and affective performances in MRIPS are indicative of learners’ performance in the real-world scenario. This content validation of MRIPS laid the groundwork for studying learning in MRIPS. As skill in MRIPS reflects real-world skill, we should expect skills learned in MRIPS to transfer to the real-world.

2. **Technological innovation:** Novel real-time and post-experiential feedback of learners’ cognitive, psychomotor, and affective performance was integrated into MRIPS. This type of feedback is known to be required for learning in interpersonal scenarios [12].

**Evaluation:** User studies were conducted to demonstrate that this feedback positively impacts learners’ cognitive, psychomotor, and affective performance in MRIPS. Because performance in MRIPS has been shown to be indicative of performance in the real-world, skills improvement due to feedback should also result in improvement in the real-world scenario.

3. **Evaluation of learning:** Learning in interpersonal scenarios takes place as a result of a repeated cycle of practice and feedback [12]. To evaluate whether learning takes place in MRIPS, we conducted a user study in which learners underwent a treatment of three MRIPS experiences each including real-time and post-experiential feedback. Before (pre-test) and after (post-test) this treatment, the learners were evaluated in the real-world scenario being simulated. Improvement from pre-test to post-test is taken as indication of learning and training transfer.

1.4.1 **Technological Innovation of MRIPS and Application to Medical Interpersonal Scenarios**

MRIPS-CBE, a simulation of the clinical breast examination was developed. A haptic interface instrumented with force sensors and cameras captures the user’s touching of the virtual human and other physical objects. This touch is incorporated into a virtual human simulation allowing for performance of the cognitive, psychomotor, and affective components of clinical breast examination on the virtual human.

MRIPS-NEURO, a simulation of a neurological examination with abnormal physical findings was then developed. A Nintendo Wii-Remote augmented with external sensing in the form of six-degree-of-freedom pose tracking provided a haptic interface for touch, hand-gestures, and hand-held tool use. This interface provided the shape,
feel, and correct kinesthetic information of real hand-held tools used in the neurological exam, as well as vibratory force feedback to indicate touching of the virtual human’s face. The augmented Wii-Remote also provides a substitute for the user’s hand, affording a robust (high update rate, low error) gesture interface.

1.4.2 Evaluation of the Validity of MRIPS for Practicing and Evaluating Cognitive, Psychomotor, and Affective Task Components

We conducted two observational studies to determine whether MRIPS-CBE elicited real-world cognitive, psychomotor, and affective skill performance. The first study demonstrated equivalent performance with MRIPS-CBE and an SP. The second study demonstrated that the learner’s prior experience with CBE of human patients impacts performance in MRIPS. These studies established the validity of MRIPS-CBE for practicing and evaluating the three skill sets of CBE.

1.4.3 Incorporation or Real-time Feedback of Cognitive, Psychomotor, and Affective Task Performance

We were motivated to incorporate real-time feedback into MRIPS by the knowledge that the learning process is driven by feedback of performance [12] and the finding that immediate, specific, non-judgmental feedback is the most important motivator for sustained learning in medical education and CBE [61][62]. We expect that incorporation of real-time, objective feedback from quantitative measures of task performance will thus improve the learning potential of MRIPS.

Feedback was created to specifically target the cognitive, psychomotor, and affective task components of the clinical breast and cranial nerve exams (summarized in Table 1-3).

For MRIPS-CBE, feedback consists of:
Feedback of the completeness and correctness of the exam are given in the form of two visualizations which address cognitive and psychomotor components of the exam. The touch map visualizes the amount of tissue palpated (and amount of tissue remaining to palpated) as well as the correctness of palpation pressure. The pattern-of-search map aids in recall of the correct pattern-of-search and visualizes how well the user is following this pattern (Figure 4-2).

A procedural “checklist” lists the important topics to ask about in the medical history and displays icons representing the positions the patient should assume during visual inspection of the breasts and the areas of the breast that should be palpated during the physical exam (Figure 4-3). This feedback is targeted to help novice learners perform more complete medical histories and exams.

Thought bubbles appearing next to the patient’s head provide feedback of how the user’s verbal and nonverbal behaviors are affecting the patient’s emotions (comfort, fear, and attitude towards the user). This feedback is targeted to guide the user to better recognize how his actions affect a patient’s feelings and when and how to better express empathy.

MRIPS-NEURO incorporated the thought bubble feedback as well as two scenario specific visualizations:

- The “H-map” visualizes the “H” pattern used to assess whether a patient has a full range of eye movement. In this assessment, the doctor sweeps his finger or a light in the pattern of an uppercase “H.” The “H-map” visualizes this pattern, the shape of which is dependent on the depth of the doctor’s finger from the patient’s head. This visualization is targeted to aid novice learners in assessing the extremes of the patient’s vision. If an incorrect pattern is used, the extremes of the patient’s vision are not adequately assessed, which may lead to incorrect diagnosis.

- The patient-vision feedback is a simulation of what a person with the cranial nerve disorder sees. By wearing an HMD, the user is able to see the virtual world through the patient’s eyes and experience the double vision and incomplete range of eye movement experienced by the patient. This feedback is targeted to aid the cognitive task of diagnosis and the affective task of perspective taking.

Evaluation of the feedback’s impact on learner behavior is provided by the evaluation of learning and training transfer (for MRIPS-CBE) and a separate user study that directly evaluated the impact of each feedback component of MRIPS-NEURO.

Results of this study indicated that the patient vision feedback improved cognitive and
affective performance, and that the H-Map feedback improved the efficiency of learners’ eye movement tests.

1.4.4 Evaluation of Learning and Training Transfer

We designed a user study, Study MRIPS-Learning to evaluate learning in users of MRIPS, and transfer of skills to a real-world interpersonal scenario. MRIPS-CBE was chosen for this study, as historical data from students learning with traditional approaches was available for comparison.

Beginning 3rd-year medical students with no prior hands-on experience with CBE were chosen as the study participants, as these participants were about to start a women’s health clerkship involving performing CBES of real patients. MRIPS-CBE is targeted for inclusion in the curriculum before this clerkship.

The study procedure was as follows: A baseline of participant performance in cognitive, psychomotor, and affective tasks was obtained by having participants perform a CBE of a standardized human patient (SP). Participants then performed three CBES in MRIPS-CBE in which they received real-time feedback of cognitive, psychomotor, and affective performance. Post-experiential feedback of performance in these three skill sets was also provided after each MRIPS-CBE interaction. Improvement was evaluated with another CBE of an SP.

We chose multiple MRIPS interactions because learning requires repeated experience, feedback and reflection [12]. Three interactions were chosen because this produced a MRIPS “curriculum” lasting three weeks, which is of reasonable length to incorporate into a medical school curriculum (in which rotations or clerkships typically last 2-3 weeks). The MRIPS interactions were spaced one week apart as this is standard in studies of learning from repeated treatments [63] and if integrated into a
curriculum would allow educators or curriculum overseers adequate time to review a student’s performance and intervene (e.g. for remediation) if necessary before the next interaction.

MRIPS provided automated evaluation of performance through data collected by the haptic and speech interfaces. This was augmented by video review by the experimenter. Expert clinicians and medical educators reviewed video of the SP interactions to evaluate participant performance with the SPs. Learning and training transfer was evaluated by analyzing changes in performance from the baseline SP to the evaluation SP and throughout the three MRIPS interactions. To evaluate the impact of real-time feedback on performance, participants’ performance in the first MRIPS-CBE interaction was compared to historical control groups of both novices and experts performing CBEs in MRIPS without real-time feedback. Historical control groups of novice medical students’ CBEs of SPs allowed us to investigate the impact of repetitive MRIPS practice and eliminate the baseline SP interaction as a source of improved performance.

For many cognitive, psychomotor, and affective tasks, participants improved significantly throughout the three MRIPS interactions. For other tasks, incremental, but not significant, improvement was demonstrated. However, participants encountered a ceiling effect on three cognitive and psychomotor tasks, likely due to the real-time feedback. Generally, participants improved their performance concurrently in all three skill sets during the three practice opportunities afforded by MRIPS-CBE. Training transfer in the form of improvement from the SP baseline to the SP evaluation CBE was demonstrated for tasks in all three skill sets.
Results of this study also demonstrated significant positive impact of real-time feedback on learner performance. In cognitive and psychomotor tasks, participants in Study MRIPS-Learning significantly outperformed those in the historical control group not receiving real-time feedback. In particular, Study MRIPS-Learning participants receiving real-time feedback of the completeness and correctness of their palpation performed CBEs with expert-level or greater completeness and use of correct palpation pressure. Learning from repetitive use of MRIPS was less clear.

Finally, we investigate the impact of having multiple MRIPS-CBE practice opportunities (vs. only one) and the impact of the baseline SP interaction on skills improvement. Compared to one MRIPS-CBE practice opportunity, participants receiving three MRIPS-CBE practice opportunities performed significantly better in cognitive and affective components of a CBE of an SP and non-significantly better in psychomotor components. After four weeks of no CBE learning opportunities (approximately the length of Study MRIPS-Learning) any improvements in performance from an SP interaction were no longer retained, demonstrating that assessing participants’ baseline skills with a CBE of an SP does not contribute to improved performance in the post-treatment evaluation with an SP. Instead, all improvements in performance from baseline to evaluation SP interactions were due to learning during MRIPS-CBE interactions and retention of the skills learned in MRIPS-CBE.

This study demonstrated that the incorporation of haptic interfaces and sensors to enable simulation of cognitive, psychomotor, and affective components of an interpersonal scenario – along with real-time feedback of performance in these components – results in concurrent learning of all three skill sets in users of the
interpersonal simulation. Repeated use of the MRIPS elicits skills improvement that transfers to the real-world interpersonal scenario. This demonstrates that an interpersonal simulation can make the leap from laboratory test-bed to a deployed curriculum component to aid in training the next generation of doctors, soldiers, and other professionals who must perform in high-stakes interpersonal scenarios.

1.5 Innovations

The innovations of this work were in designing and evaluating a novel approach to interpersonal simulation for interpersonal skills training. We designed haptic interfaces instrumented with physical sensing, e.g. force sensors and optical tracking, to afford touch between a human and a life-sized virtual human agent. These interfaces are the first haptic interfaces to a life-sized virtual human.

Bidirectional (human to virtual human and virtual human to human) touch is incorporated into the virtual human’s simulation as a means of performing the cognitive, psychomotor, affective, and compound cognitive-psychomotor tasks of the interpersonal scenario. In its cognitive and affective role, touch provides a means for communicating with the virtual human and for the virtual human to communicate with the user. This is the first incorporation of touch as a means for communication with a virtual human. The haptic interfaces afford psychomotor and cognitive-psychomotor task performance that involves touching of the virtual human and manipulation of hand-held tools. The interfaces are designed in a manner that provides kinesthetic feedback consistent with the of the real-world scenario being simulated – i.e. using these haptic interfaces, one can learn the muscle movements to perform the psychomotor task in the real-world scenario.
We also designed novel real-time, objective, dynamically generated, in-situ presented feedback of quantitatively measured learner performance. This performance feedback was designed to train psychomotor, cognitive, affective, and cognitive psychomotor task components. In user studies of MRIPS-CBE and MRIPS-NEURO, cognitive, psychomotor, and affective feedback improved user performance in those three skill sets. These feedback methods also provide educators with quantitative measurements of user performance in psychomotor, affective, and joint cognitive-psychomotor task components, uniquely allowing educators to automatically, and more precisely evaluate learner performance.

Applied to simulating CBE and neurological examination, MRIPS is the first simulation of medical physical examination which targets the visual-kinesthetic learning style favored by a majority of medical students, while also simulating both interpersonal and physical exam components of the exam scenarios.

Finally, we provide evidence that learning occurs in users of interpersonal simulation and that learned skills transfer to the real-world scenario being simulated. Although our study of learning was confined to the CBE scenario and should be considered only a first-pass at evaluating learning and training transfer, it represents a comprehensive attempt to determine whether interpersonal simulation is ready for deployment into real-world interpersonal curricula.
Table 1-1. Affordances and drawbacks of prior educational approaches and MRIPS for teaching cognitive, psychomotor, and affective aspects of high-stakes interpersonal scenarios

<table>
<thead>
<tr>
<th>Approach</th>
<th>Cognitive</th>
<th>Psychomotor</th>
<th>Affective</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer</td>
<td>Yes</td>
<td>Yes, for a limited set of scenarios: “there is no place for peer genital, rectal, or female breast exams in the curriculum” [17]</td>
<td>No, not taken seriously by ~33% of students [8]</td>
<td>Can not be used in high-stakes intimate exams, limiting the set of applicable scenarios [6]. No abnormal findings and limited cultural differences [19].</td>
</tr>
<tr>
<td>Human actor [16]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited availability and set of scenarios [6]. No abnormal findings [7].</td>
</tr>
<tr>
<td>IPS [10]</td>
<td>Yes</td>
<td>Yes, limited to pointing and iconic gestures or encumbered by gloves or body suits [10][11][64]</td>
<td>Yes, if taken seriously *</td>
<td>Psychomotor tasks limited to pointing and iconic gestures. No touch [10][11].</td>
</tr>
<tr>
<td>MRIPS</td>
<td>Yes (evaluated in Chapter 5)</td>
<td>Yes (evaluated in Chapter 5)</td>
<td>Yes, if taken seriously (evaluated in Chapter 5) **</td>
<td>~10% do not approach seriously. ** Similar drawbacks as IPS with regard to robustness of verbal interaction (Section 2.2.1).</td>
</tr>
</tbody>
</table>

(*) In Section 3.4 we present the results of a pilot study which demonstrated that learners’ affective behaviors in MRIPS were more similar to a human patient interaction than were learners’ affective behaviors in IPS [82]. (**) Section 4.2 presents evidence that if treated seriously, learners’ affective performance in MRIPs is equivalent to that with human actors. However throughout the evaluations of MRIPS, we have found that a small percentage (<10%) do not approach MRIPS seriously.
### Table 1-2. Traditional and simulation approaches to teaching intimate exams

<table>
<thead>
<tr>
<th>Approach</th>
<th>Interpersonal aspects simulated?</th>
<th>Exposure to abnormal findings?</th>
<th>Feedback?</th>
<th>Learning style emphasized:</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPs [16]</td>
<td>Yes</td>
<td>No</td>
<td>Real-time: affective, performance (social, auditory) Post-experience: completeness of interview and exam (auditory)</td>
<td>Kinesthetic</td>
</tr>
<tr>
<td>Expert observer [12]</td>
<td>Yes</td>
<td>By chance (&quot;catch as catch can&quot;)</td>
<td>Real-time: correctness of exam technique (auditory)</td>
<td>Kinesthetic, auditory [42]</td>
</tr>
<tr>
<td>Physical sim [65][66]</td>
<td>No</td>
<td>Yes</td>
<td>Only if used in peer group learning. Real-time: correctness of exam technique (auditory)</td>
<td>Kinesthetic</td>
</tr>
<tr>
<td>Hybrid (SP + physical sim) [7]</td>
<td>Yes</td>
<td>Yes</td>
<td>Real-time: affective performance (social) Post-experience: completeness of interview and exam (auditory)</td>
<td>Kinesthetic</td>
</tr>
<tr>
<td>Physical sim + sensing [39]</td>
<td>No</td>
<td>Yes</td>
<td>Real-time: charts and diagrams (visual)</td>
<td>Kinesthetic, visual</td>
</tr>
<tr>
<td>MRIPS</td>
<td>Yes</td>
<td>Yes</td>
<td>Real-time and post-experience: affective performance, completeness and correctness of exam and interview (social, auditory, visual)</td>
<td>Kinesthetic, visual</td>
</tr>
</tbody>
</table>

### Table 1-3. Feedback in MRIPS

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Cognitive</th>
<th>Psychomotor</th>
<th>Affective</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRIPS-CBE</td>
<td>Touch map (coverage)</td>
<td>Touch map (pressure)</td>
<td>Thought bubbles</td>
</tr>
<tr>
<td></td>
<td>Pattern-of-search map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Procedural checklist (breast history, visual inspection)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRIPS-NEURO</td>
<td>Procedural checklist</td>
<td>H-map</td>
<td>Thought bubbles Patient vision</td>
</tr>
<tr>
<td></td>
<td>H-map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patient vision</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1-1. The simulator of Pugh et al. [38] provides feedback of pressure as a series of meters representing the value at each of the 11 sensors, and shows coverage as lit/unlit dots representing whether each sensor has been touched.
CHAPTER 2
REVIEW OF LITERATURE

The goals of this chapter are to impart an understanding of prior approaches to interpersonal simulation and to motivate MRIPS’ incorporation of touch to enhance the capabilities of interpersonal simulation to train cognitive, psychomotor, and affective tasks. Prior work relative to our goal of educating interpersonal scenarios through the use of simulation is reviewed. The review of prior work related to the motivation and implementation of specific components of this dissertation is left to these components’ respective chapters.

In this chapter, Section 2.1 investigates a theoretical foundation for the use of virtual humans for interpersonal simulation. This prior work demonstrated social responses to virtual humans and investigated behavioral changes resulting from social interaction with virtual humans.

Section 2.2 details current approaches to interpersonal simulation with virtual humans, connecting these previous approaches to the novel approach of MRIPS.

Section 2.3 motivates the incorporation of touch into interpersonal simulation by detailing the many uses of touch in real-world interpersonal scenarios.

As we have applied MRIPS to medical physical examination interpersonal scenarios, it is useful to be aware of related approaches other than virtual humans. Section 2.4 provides a brief overview of previous and contemporary approaches to simulating physical examination.

2.1 Foundations for Interpersonal Simulation with Virtual Humans

This section investigates a theoretical foundation for the use of virtual humans for interpersonal simulation. This prior work demonstrated social responses to virtual
humans and investigated behavioral changes resulting from social interaction with virtual humans.

2.1.1 Social Responses to Virtual Humans

The goal of interpersonal simulation is for a virtual human to effectively simulate a human interaction partner. This goal can not be achieved without first demonstrating that virtual humans are treated as social entities. The prior work detailed in this section has accomplished this, demonstrating that virtual humans elicit social responses that are consistent with human-human interaction in a variety of real-world scenarios.

A field which has extensively used virtual humans to provoke social responses is the field of virtual reality exposure therapy (VRET). VRET provides exposure therapy (used to conquer phobias, post-traumatic stress, and other psychological disorders) in a virtual world, allowing the clinician to standardize and control stimuli to which the patient is exposed. In VRET, real-world social fears have been elicited by virtual worlds inhabited by virtual humans, e.g. a virtual audience elicits anxiety in users with a fear of public speaking [67]. The presence of virtual humans also provided the social stimuli to increase cravings for smoking among users of a virtual environment to treat nicotine addiction [68]. A “bar” virtual environment inhabited by virtual humans who directed their attention towards the user elicited anxiety in socially phobic users [69].

Other real-world social responses to VHS include social inhibition when being observed by a virtual human during a complex task [70] and affording a virtual human a similar amount of personal space as afforded to a human [71]. Social interactions with virtual humans also elicit behavioral expression of users’ implicit biases, such as skin-tone bias [20].
It is important to note that these social responses were elicited without extensive communication between human and virtual human. In much of the prior research, the user’s actions were not able to impact the virtual human’s behaviors (i.e. the virtual human simulation accepted no input from the user). Instead it was the virtual human’s appearance or unidirectional verbal and nonverbal communication that impacted the user’s attitudes and actions. This differs from interpersonal simulation, in which verbal and nonverbal interaction between human and virtual human is the focal point.

2.1.2 Toward Changing Human Behavior

The goal of MRIPS is to improve a learner’s cognitive, psychomotor, and affective behavior in an interpersonal scenario. The prior work covered in this section has shown that human behavior can be altered by interacting with a virtual human. However, this prior work has stopped short of demonstrating learning of real-world social scenarios as a result of using interpersonal simulation.

Human opinions are known to be altered by the expression of the opinions of their social interaction partners. This phenomenon was demonstrated for virtual human social interaction partners as well. Zanbaka et al. found that virtual humans presenting an unfavorable argument, e.g. arguing for tuition increases, were able to persuade their human interaction partners to change their attitudes towards the topic [72].

Further work has shown that a person’s attitudes towards a subject can be changed by an interaction with a virtual human. In an effort to pique middle school girls' interest in engineering, an interaction with a female virtual human engineer changed students’ attitudes towards engineering. Conversing with the virtual human increased students’ interest in engineering and performance in math. This was attributed to improving students' belief in their abilities [73].
Beyond the altering of expressed attitudes, it has been shown that in some scenarios, social interaction with virtual humans can improve human social behaviors. VRET incorporating virtual humans as observers has improved patients’ management of their fears in the real-world scenario being simulated. Repeated delivery of a speech to a virtual human audience, along with anxiety management treatment from a human expert, resulted in decreased anxiety in later speeches to human audiences. This work demonstrates that these virtual experiences can alter behavior. However it did not directly demonstrate training of a skill, as an increase in speech quality was not measured [74]. The decrease in anxiety demonstrates a positive impact of a virtual experience on user behavior, but was partially due to conditioning and use of breathing exercises. Improvement in the real-world scenario was not demonstrated to be a result of interpersonal skills learned in the simulation.

Interaction with virtual humans has also been targeted towards teaching psychomotor skills. Users unfamiliar with South Indian greeting protocols performed greetings more correctly after watching a virtual human perform the greeting than after reading a text-based instructional booklet [64]. This work stopped short of demonstrating short-term learning by not measuring an improvement in participants’ greetings from before to after viewing the virtual human greeting.

A similar approach was taken to improving children’s behavior in the cognitive task of determining a safe gap in traffic for crossing an intersection on a bicycle. Children’s bicycle riding behavior was influenced by a virtual peer. When riding with a virtual peer who made unsafe crossings through traffic, the child mimicked these unsafe behaviors [75]. Although the authors’ assumed this would persist as a change in behavior in real-
world bicycling, the study conducted was not designed to measure learning or behavioral changes.

This previous work has shown that social interactions with virtual humans have the potential to alter human behavior, at least in the short-term and in the presence of the virtual human. Our work goes beyond this to investigate improvement in specific real-world interpersonal scenarios as a result of practicing these interpersonal scenarios with a virtual human.

2.2 Interpersonal Simulation with Virtual Humans

MRIPS expands on prior approaches to interpersonal simulation by incorporating touch as a means to perform cognitive, psychomotor, and affective components of interpersonal scenarios. This section describes the prior approaches to interpersonal simulation and aspects of these approaches that MRIPS expands upon.

2.2.1 Current Approaches to Interpersonal Simulation

We have defined interpersonal simulation as the recreation of a real-world interpersonal scenario which replaces a human with a virtual human confederate. The primary goal of interpersonal simulation has been to train users for corresponding real-world interpersonal scenarios.

For this reason, many current approaches to interpersonal simulation have strived to provide natural interfaces to interact with the virtual human. Typically the virtual human is presented at life-size in a head-mounted display (HMD) or on a large screen or projected display. Users are able to communicate with the virtual human through natural speech, which is processed by speech recognition software and the resulting text matched to a database of phrases understood by the virtual human. When the user
utters a phrase in this database, the virtual human replies with speech, either pre-
recorded or text-to-speech, and gestures and facial expressions [10][13][15][21][31][33].

The primary communication modality in these simulations is speech. Speech
recognition rates vary from user to user (e.g. due to accents and speaking style – long
utterances and mumbling are two qualities which result in poor speech recognition).
Additionally, speech matching rates vary based on the matching approach, e.g. 60-70%
[15], 53-62% [76], 75% [77]. These figures may include a significant number of false
positives (incorrect matches which trigger virtual human responses which appear to the
user to be “on-topic” but may not convey the same information as the correct response).
Artstein et al. report 33% false positives using the matching approach of Leuski et al.
[76][78]. These interpersonal simulations are meant to be autonomous, and their
imperfect speech interfaces negatively impact the usability of these systems in an
autonomous fashion (i.e. the user might get stuck because the virtual human can not
understand his question). For these reasons interpersonal simulations which rely
primarily or solely on speech as the means for communication with the virtual human
are often augmented by a human controller to evaluate the potential of the system to
simulate a specific scenario. This behind-the-scenes “wizard-of-oz”, is able to manually
trigger virtual human responses if speech recognition or matching fails, e.g. [14]. In this
work, we used a wizard-of-oz to augment speech recognition when investigating
whether MRIPS has the potential to elicit real-world psychomotor, cognitive, and
affective behaviors (Chapter 5). In these user studies, the wizard-of-oz triggered the
virtual human’s response if speech recognition failed for a user’s question. However,
when evaluating MRIPS for learning and training transfer, verbal communication relied
solely on the automated speech matching approach of Johnsen et al. [15]. With respect
to the efficacy of speech interfaces, we have also observed that the incorporation of
touch and tool-use as additional interaction modalities appears to mitigate user
frustration in the presence of poorly performing speech recognition and matching.

A small number of interpersonal simulation approaches allow the user to
communicate nonverbally with the virtual human using a small set of simple gestures.
Users of a medical interview simulation wore a glove augmented with optically tracked
fiducials on their dominant hand. This glove allowed the user to point to a spot on the
virtual human’s body and ask “does it hurt here?” [10]. Cassell et al. used position data
of the user’s hands and head, captured using unencumbering vision-based tracking, as
inputs to Rea, a virtual human real estate agent. Gesture input was limited to detecting
when the user was moving his or her hands, indicating that the user wanted to speak
[11]. In an evolution of this system, Gandalf, a cartoonish virtual character who served
to guide users through a virtual solar system, was able to recognize iconic gestures,
e.g. holding a hand up to signal “stop”, in addition to pointing, by having users wear a
suit incorporating orientation and position sensors on the hands and chest [79].
Although not used as inputs to the simulation, users who practiced cultural greetings
with a virtual human wore position-orientation sensors on the hands, waist, and head to
record their body movements during the greeting [64].

These prior approaches to incorporating nonverbal communication into
interpersonal simulations have primarily relied on encumbering tracking devices (e.g.
gloves, wires, headbands, body suits) and have enabled only simple iconic gestures to
be input to the virtual human. My work incorporates unencumbered, robustly-tracked
haptic interfaces for hand-held tool manipulation for hand-gestures and accomplishing psychomotor tasks not involving touch.

Universally, prior interpersonal simulators lack a natural means of touching for hands-on psychomotor interaction. For this reason, interpersonal simulation applied to scenarios requiring psychomotor tasks (with or without touching) has relied on unnatural keyboard-and-mouse interfaces.

To apply interpersonal simulation to scenarios requiring tool use, touch, or complex gestures, non-natural mouse-keyboard interfaces were used in addition to the natural speech interface. This approach has been applied to scenarios such as clinical examination and battlefield operations [30]. The portion of the scenario requiring touching or other psychomotor tasks is conducted separately from the portion of the scenario involving only verbal interaction. This separation of components by interaction modality prevents touching from being a mechanism for nonverbal communication with the virtual human and makes the simulation dissimilar from the real-world scenario. Additionally, use of the unnatural mouse-keyboard interface for touch negates the possibility of training psychomotor components of these interpersonal scenarios, as different muscle movements are used with the mouse and keyboard than are used in the real-world scenario.

Expanding on the affordances of these prior approaches to interpersonal simulation, MRIPS incorporates natural interfaces for verbal, gestural, and touch interaction with a life-sized virtual human. The incorporation of haptic interfaces and sensing of user manipulation of these interfaces affords concurrent use of speech,
gestures, and touch as means for communication (cognitive and affective) and performance of psychomotor tasks required of the interpersonal scenario.

### 2.2.2 Existing Interpersonal Simulators

This section describes the application of prior interpersonal simulation approaches to the fields of military, law enforcement, and medicine. Evaluation of these simulators has typically been restricted to face validity or usability, though the medical history simulator of Johnsen et al. has been validated for evaluating medical students’ cognitive history taking skills [81].

Within the military domain, interpersonal simulators have been employed to train end-users’ communication skills in military crisis management [13] and negotiation [33] scenarios. The simulation which addressed these scenarios is the Mission Rehearsal Exercise system (MRE) developed by the University of Southern California Institute for Creative Technologies (ICT). The MRE allows a user to communicate verbally, using natural speech, with virtual humans playing the roles of allied soldiers and foreign civilians.

Bridging military and medical domains, USC ICT has also developed male and female virtual humans, Justin and Justina, who simulate patients with conduct disorder (Justin) and post-traumatic stress disorder (Justina). Preliminary testing with mental health clinicians have established face validity of this interpersonal simulation to provide novice clinicians exposure to these mental health conditions [31][32].

Within the medical domain, Johnsen et al. created IPS (The Interpersonal Simulator). The goal of IPS is to train health profession students’ communication skills in medical interviews [10]. In IPS, the user converses with a life-sized virtual human patient who responds to user speech with pre-recorded human speech, gestures, and
facial expressions. Additional nonverbal input was provided through a simple gesture interface limited to recognizing pointing [10]. IPS does not provide any mechanism for touching the virtual human. Users of this simulation treated the virtual human patient similarly to a standardized human patient, but were not sincere in their use of empathy and had difficulty achieving rapport [80]. However, this interpersonal simulation was validated for evaluating novice medical students’ interview skills, as students’ performance with the virtual human predicted their performance with a standardized human patient [81].

The speech recognition and matching capabilities and virtual human appearance of MRIPS are directly based on the simulation of Johnsen et al. [81]. With the incorporation of touch as a means for communication, we have observed improvement in the frequency of users’ empathy towards the virtual human [82].

Also in the medical domain, smaller than life-sized virtual humans were incorporated in simulations for practicing clinical examination skills [30] and informed consent interviews [29]. These systems used a natural speech interface in addition to a mouse-and-keyboard interface for performing nonverbal components of the scenarios. The same approach was taken to simulating police interviews of mentally ill suspects [22].

While these systems have received various degrees of evaluation, none have been evaluated for learning or improvement in the real-world scenario being simulated. In this dissertation, we take this next step, evaluating what (if any) learning occurs in MRIPS and whether skills learned in MRIPS transfer to the real-world.
2.2.3 Incorporation of Feedback in Interpersonal Simulation

In real-world interpersonal scenarios, receiving feedback during and after one’s performance is necessary for learning [12][61][62]. Interpersonal simulations are thus beginning to be augmented with real-time and post-experiential (i.e. After Action Review) feedback. MRIPS also incorporates real-time and post-experiential feedback of a learner’s performance. The feedback provided in MRIPS draws on the work described in this section to innovate new approaches to providing feedback of cognitive, psychomotor, and affective components of learner performance.

In the Virtual Environment Cultural Training for Operational Readiness (VECTOR) simulation by CHI Systems Inc, soldiers negotiating with virtual Iraqi civilians receive feedback of the virtual human’s emotional state in real-time, displayed textually, e.g. “Neutral” or “Anxious” appears over the character’s head and changes depending on the user’s actions [21]. A similar approach was taken with USC ICT’s ELECT BiLAT, also a cultural competency training system. The virtual human’s level of trust of the user is indicated by a one-dimensional meter, similar to a completion or “loading” meter [83]. No evaluation of this feedback was reported. This feedback guided the user’s affective and cognitive performance, and is the only instance of real-time feedback of affective performance we have come across in interpersonal simulation. MRIPS takes a different approach to affective guidance. Instead of explicitly stating how well the learner is performing (e.g. “50% trust”), the virtual human emits “thought bubbles” indicating how the quality of the user’s affective performance might affect a real patient’s emotions and attitudes.

To provide post-experiential feedback of cognitive performance, Raij et al. incorporated an After-Action Review system into the interpersonal simulation of
Johnsen et al. for practicing medical history taking [10]. This system allowed users to receive feedback in a self-guided manner, navigating a traditional monitor-and-keyboard interface which provided video of the user’s performance, expert performance, graphs of how users progressed through topics in the medical history, and visualization of the user’s head gaze at the virtual human patient. The system aided users in reflecting on their performance, exhibited by changes in self-ratings of performance, and may have motivated users to improve their behavior in future interviews, as users reported they would change, but this was not explored with further user studies [84].

Expanding this approach to affective feedback of performance, Raij and myself built a post-experiential feedback system which allowed medical students to relive their breast exam of a virtual human – through the eyes and body of the virtual human [85]. This experience had the goal of improving medical students’ empathic behavior and perspective taking (the affective components of the medical interview). The system created by Raij et al. serves as inspiration for a novel feedback experience designed to enhance both cognitive and affective performance in MRIPS (Chapter 8).

2.3 Motivation for Touch in Interpersonal Simulation

One of the primary innovations of MRIPS is to incorporate touch as a means of communication between human and virtual human. We are motivated to incorporate touch into interpersonal simulation due to the important and widespread roles touch plays in communication between two humans. Interpersonal touch is a critical component of communication in real-world interpersonal scenarios; thus, without affording touch, current interpersonal simulation provides an incomplete simulation of a real-world interpersonal scenario. Additionally, the incorporation of touch into both non-inhabited (no virtual humans) and collaborative (multiple non-colocated human users
with avatars) virtual worlds has been shown to enhance these virtual worlds’ believability and usability [86].

2.3.1 The Role of Touch in Communication

In clinical, professional, and casual social situations, touch between two humans is used to communicate disagreement, agreement, appreciation, interest, intent, understanding, affection, caring, support and comforting [35][87]. Bidirectional touch between interaction partners elicits more positive attitudes towards the interaction partners than communication through speech and gestures alone [88]. Even an unexpected touch from a stranger can elicit positive affect towards the stranger [34].

Touch is essential in many social interactions, such as those between medical doctors and patients. Caregivers touch patients to communicate reassurance and empathy, convey the idea that the caregiver is helping the patient, and achieve patient compliance, e.g. following of treatment plans [36][89]. Touching the patient improves doctor-patient communication by increasing patient verbalization, self-disclosure, and rapport, and producing positive attitudes towards the caregiver [90][91]. Touch from patient to caregiver is used to communicate friendliness, gratefulness, and establish a positive patient-doctor relationship [92].

By incorporating touch into interpersonal simulation, not only is performance of psychomotor task components enabled, but communication between the human and virtual human is made more like communication between two humans. In evaluations of MRIPS, we have observed users touching the virtual human to communicate many of these constructs, e.g. empathy, concern and comforting; instruction and compliance; and social norms such as touch in greeting [93].
2.3.2 Touch in Virtual Environments

Prior work has demonstrated that adding touch to uninhabited virtual environments and collaborative (inhabited by other humans’ avatars) virtual environments increases the realism and usability of these virtual environments. This provides a theoretical background against which we have made observations that, with the addition of touch, MRIPS users treat the MRIPS virtual human more similarly to a real human than has been observed in prior interpersonal simulations lacking touch [82].

Adding haptics to virtual environments has been shown to increase the believability of the experience. The addition of passive haptics to the UNC Pit environment, a stress invoking walk over a virtual chasm, increased users’ sense of presence, i.e. being there [144]. Haptics also increases the sense of co-presence, i.e. being with another, in collaborative virtual environments. A passive-haptic “lazy susan” increased remote users’ sense of being co-located around a virtual table [94]. Remote users collaborated in a shared Unified Modeling Language editor more effectively when using active-haptic cursors provided by a PHANTOM Omni [95]. Remote users collaborating to move a ring along a curved wire in a virtual world reported a higher sense of togetherness when active-haptic feedback was given than when only visual feedback was given [86].

Haptic interfaces have also been used to interact with a virtual human in game-like scenarios. An active-haptic interface allowed a user to play catch with a virtual human [145]. The passive-haptic interface of a real checkers set allowed a human to play a game of checkers with a virtual human [96].

Prior to MRIPS, the work that had come the closest to enabling touch as a means for communication with a virtual human is that of virtual interpersonal touch. Bailenson
and Yee proposed virtual interpersonal touch as the touching of a smaller than life-size virtual human using an active-haptic interface. An active-haptic interface is one that mechanically applies force to the user, as the user manipulates the interface elements. A study found that participants who used a Phantom Omni force-feedback device to clean virtual dirt off of a virtual human’s body touched female virtual humans with less force than male virtual humans. This result fit with known results from psychology literature concerning gender effects on interpersonal touch between two humans. However, the virtual human did not communicate with the user or react to the touch. It is important to note that this was not considered to be an interpersonal touch between a human and virtual human, as the cleaning was not a type of social touch and there was no communication between the human and virtual human [97]. Bailenson and Yee’s approach of affording touch using currently available active-haptic devices to provide a realistic feel of touch would be logistically and mechanically difficult to extend to touching of a life-sized virtual human. For this reason, MRIPS takes the approach of using passive haptics when a large area of the life-sized virtual human’s body is to be touched by the user. A passive haptic interface does not mechanically apply force to the user, but provides the user feedback through the interface elements’ shape, weight, and texture.

Being touched has yet to be explored in interpersonal simulation or social virtual worlds. However, it has been shown that purely virtual stimuli can elicit feelings of being physically touched. These visual touches or pseudo-haptics were explored by Biocca et al. Participants who manipulated a virtual spring, the visual analog of a physical force, reported sensing a haptic resistance [98]. Pusch et al. found that an
illusory haptic sensation was achieved by visually displacing the user’s hand when placed into a virtual force-field [99].

My work expands on prior research into the impact of touch by using passive and active-haptic interfaces to afford bi-directional interpersonal touch between human and virtual human, providing touch as an additional modality for performing the cognitive, psychomotor, and affective components of interpersonal scenarios.

2.4 Other Approaches to Medical Interpersonal Simulation

We have targeted MRIPS to simulate medical physical examination scenarios which are underserved in terms of practice opportunities, standardization, and feedback [9]. Approaches other than virtual humans have been previously used to provide additional practice in medical procedures. These approaches have incorporated computer based simulation but are predominantly physical (i.e. mechanical) approaches.

One of the most mature and sophisticated simulation approaches for learning medical procedures is the human patient simulator (HPS). The HPS is a mannequin simulator that provides a realistic, full-sized human shaped haptic interface to an underlying agent controlling the patient’s physiological state. The mannequin can breathe, blink, and can be intubated, anesthetized, injected with medication, and can have his blood pressure and heart rate measured. The healthcare provider can communicate verbally with the HPS. However, communication is not agent controlled. Instead, a “wizard-of-oz” human controller “listens” and “speaks” for the HPS patient. For this reason, the HPS is primarily used to train scenarios in which interpersonal skills are not emphasized, e.g. procedures in which the patient has been anesthetized. Additionally, though realistic, feedback consists only of feedback that is received during
real procedures, e.g. monitoring of vital signs [55] – the HPS does not provide feedback that is unavailable with a human patient, e.g. viewing inside the patient to see how anesthetic flows through the patient’s system. Though the haptic interface of MRIPS is not as sophisticated, MRIPS enables on-demand (completely automated) practice of interpersonal and physical exam skills, and provides feedback beyond that available in purely physical environments, by co-locating visual feedback of performance with the anatomy being examined.

Fledgling approaches to medical simulation using robotics are in development. Robotics techniques have recently been applied to create human-form robot patients capable of verbal communication [100] and realistic movement. Nonverbal communication of this approach appears limited. The current version of the robot patient’s nonverbal gestures consist only of exhibiting symptoms. Touch input to the simulation is limited to triggering 1-2 sensors placed at the appendix. Using touch for communication with this approach is certainly possible but has not yet been demonstrated.

These predominantly physical approaches encounter difficulties in presenting feedback, and simulating abnormal findings and cultural differences. By taking a mixed reality approach that merges the flexibility of virtual human agents for simulating a wide variety of presentations and communication with the psychomotor capabilities of physical simulation, MRIPS captures the “best of both worlds.”
CHAPTER 3
MRIPS DESIGN PRINCIPLES AND DEVELOPMENT OF MRIPS-CBE

This chapter describes the design principles behind a mixed reality interpersonal simulation (MRIPS) and the development of MRIPS-CBE, an interpersonal simulation of a clinical breast examination. This simulation along with pilot study results was published in the proceedings of the IEEE Virtual Reality 2008 conference (best paper) [82]. An expanded version of this paper incorporating results of a study comparing usability and acceptability of MRIPS-CBE to standardized human patients was published in the IEEE Transactions on Visualization and Computer Graphics journal, May/June 2009 issue [101]. Implementation of bidirectional touch and related user study results were published in the proceedings of the IEEE Virtual Reality 2009 conference [93]. The treatment given in this chapter adds significant detail concerning the cognitive, psychomotor, and affective elements of MRIPS-CBE, the haptic interface implementation, and the virtual human simulation implementation.

Collaborators: The speech interface was designed by Kyle Johnsen. The virtual human simulation module of MRIPS-CBE is an expanded version of the virtual human simulation module created by Kyle Johnsen and Andrew Raij [10]. The appearance of the virtual human was created by Brent Rossen, Corey Forbus, and myself. Medical collaborators Scott Lind and Adeline Deladisma provided medical information pertaining to the clinical breast exam. The haptic interface of MRIPS-CBE is a significantly expanded version of a breast palpation trainer developed by Carla Pugh at Northwestern University.

Personal contributions: I developed the current version of the haptic interface, significantly expanded an existing virtual human simulation module, incorporating touch
inputs and outputs into the simulation, developed a framework and hardware for touch outputs (bidirectional touch), incorporated video of haptic interface manipulation into an existing rendering module, and incorporated manipulation of physical objects into the simulation.

Relevance to thesis: The thesis states that interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance improves users’ psychomotor, cognitive, and affective skills in an interpersonal scenario. This chapter describes the design of a mixed reality interpersonal simulation incorporating haptic interfaces and providing real-time feedback (Chapter 6-7). The simulation described in this chapter, MRIPS-CBE, is used to evaluate learning and training transfer in MRIPS (Chapter 9).

3.1 MRIPS Design Principles

Mixed interpersonal simulation (MRIPS) incorporates haptic interfaces augmented with real-time sensing into an interpersonal simulation – augmenting bidirectional verbal communication with bidirectional touch between human and virtual human to enhance communication and train psychomotor tasks.

The two main components of MRIPS are a virtual human simulation and a haptic interface. However, MRIPS does more than just merge the communication-oriented virtual human simulation and the psychomotor-oriented physical simulation (of the haptic interface). In addition to merging virtual humans and physical simulation, MRIPS also augments both components: the communication affordances of the virtual human simulation are augmented with touch-driven communication (e.g. touch to communicate instruction); the affective affordances of the virtual human simulation are augmented with the affective use of touch (e.g. a comforting touch); and the psychomotor
affordances of physical simulation are augmented with sensing (e.g. force sensors, cameras) to track manipulation of the haptic interface and extract data of learner performance. The result is a simulation of the cognitive, psychomotor, and affective components of an interpersonal scenario.

The merging and augmenting of affordances is paralleled by the merging and augmenting of the physical and virtual components of MRIPS.

The haptic interface and virtual human are merged through spatial registration, meaning they occupy the same volume in space. Registration conceptually merges the virtual human and haptic interface into a single entity, a mixed reality human, that exists in physical (tangible) and virtual (intangible) forms.

The concept of the mixed reality human is extended to the augmentation of physical and virtual components. The visual appearance of the mixed reality human consists of a virtual human augmented by real-time video streams of the haptic interface and the user’s manipulation of the haptic interface. The simulation controlling the mixed reality human provides verbal inputs and outputs and gestural outputs of a virtual human simulation, augmented with touch inputs and outputs. These touch inputs are fed with data read from sensors augmenting the haptic interface; touch outputs drive active-haptic components of the haptic interface.

These virtual and physical components of the MRIPS design are presented as two interfaces and a simulation which takes inputs and produces outputs at each interface: the visual interface, the physical interface, and the mixed reality human simulation. Taken as a whole, these three pieces of this design simulate the cognitive, psychomotor, affective, and joint cognitive-psychomotor components of an interpersonal
simulation. Although the affordances of the components are adapted to the specific interpersonal scenario simulated, a general set of affordances is given here:

The physical interface represents the physical embodiment of the mixed reality human. The physical interface incorporates sensing of the user’s actions and manipulations of passive-haptic and active-haptic interface elements, producing information describing the user’s touching of the mixed reality human, props, and tools. The affordances of the physical interface are touching of the mixed reality human, props, and tools for cognitive tasks (e.g. touches communicating instruction), affective tasks (e.g. comforting touches), and psychomotor tasks (e.g. physical examination). The physical interface also affords the mixed reality human’s touching of the user, through active-haptic interface elements, for purposes of communication. These affordances can be summarized as bidirectional nonverbal touch communication, human-to-virtual human nonverbal gestural communication, and psychomotor and cognitive-psychomotor task performance.

The visual interface displays the virtual human, augmentation with physical interface elements and a real-time video avatar of the user. The affordances of the visual interface are bidirectional verbal communication, and virtual human-to-human nonverbal communication (gestures and facial expressions).

The mixed reality human simulation takes inputs of touch, speech, and hand-held tool use, and produces as outputs virtual human speech, gestures, facial expressions, and touch.

The remainder of this chapter describes the application of the MRIPS design principles to the simulation of the interpersonal scenario of clinical breast exam. The
implementation of this simulator, MRIPS-CBE, begins in Section 3.3; Section 3.2 first provides background on the clinical breast exam interpersonal scenario.

### 3.2 Clinical Breast Examination

#### 3.2.1 CBE Procedure

A clinical breast exam has two components which can happen concurrently: the medical interview and the physical exam.

The medical interview is a 10-minute conversation in which the healthcare provider and patient exchange information. Each of the communication partners possesses unique goals for this interaction.

The goals of the healthcare provider are to gather key facts of the patient’s condition (e.g. principal complaint: the patient has found a hard mass in her left breast; family history: her sister had breast cancer) and to put the patient at ease and develop rapport with the patient. These goals are achieved through asking questions of, and expressing appropriate empathy to, the patient. When interacting with a human patient, interpersonal touch is critical for allowing the healthcare provider to develop rapport and express empathy [34][36][89]. Previous user studies of the interpersonal simulator of Johnsen et al. revealed that the healthcare provider had difficulty building rapport with a virtual human patient with a breast cancer fear due to the lack of interpersonal touch [10][80]. MRIPS-CBE allows the healthcare provider to communicate both verbally and through touch, and the patient to synthesize verbal and touch inputs, in order to better accomplish these goals.

The patient has two main goals: to receive information about her condition, and to be comforted. The patient accomplishes these goals by asking questions of the healthcare provider – however, the patient’s willingness to reveal information and ask
questions is dependent on the rapport achieved with the healthcare provider. Appropriate use of interpersonal touch by the provider can help in this regard [36][89][90]. MRIPS simulates the patient’s anxieties and desire to find out what may be wrong with her, by verbally challenging the healthcare provider with questions such as “do you think I could have cancer?”

The physical exam consists of a visual inspection of the patient’s breasts followed by palpation (touching). Interpersonal skills continue to be required during the exam, as the interview may continue during the physical exam and the healthcare provider should talk the patient through the procedures as they are about to be performed.

Visual inspection begins in a sitting position, as the healthcare provider asks the patient to remove her gown and assume three poses: arms relaxed, arms raised above the head, and hands pressed on hips with chest flexed. The patient is then asked to close her gown and lie down, and the healthcare provider again visually inspects the breasts in the prone position. In this inspection, the healthcare provider looks for visible abnormalities such as asymmetry, puckering, and redness.

The provider then palpates one breast at a time, with only one breast exposed at a time to minimize patient discomfort and embarrassment. MRIPS-CBE allows the learner to manipulate a physical gown in order to achieve this correct “draping” of the patient during palpation.

Though there are multiple approaches to palpating the patient’s breasts, the currently recommended approach is the Mammacare method which has been found to maximize sensitivity of the palpation to finding breast masses [66][102]. The patient lies on her back with the arm corresponding to the breast being palpated placed behind her
head. The breast is palpated in a vertical-strip pattern-of-search, i.e. a “lawnmower” pattern. Each palpation consists of circular motions at low pressure (feeling the superficial layer), medium pressure, and high pressure (pressing to the chest wall), with each palpation lasting approximately one second. Visual and verbal communication is also important during palpation, as the patient may experience pain which could be expressed verbally but may be more subtly expressed in facial expressions. The healthcare provider should recognize these expressions, and might also preface the exam by asking “let me know if you experience any tenderness.” Verbal communication from the provider should also be used to put the patient at ease and to communicate what the palpation will feel like for the patient. Additionally, if, e.g., the patient has previously found a mass, the provider may ask questions such as “is this the mass that you found?” when the mass is palpated.

After palpation is completed, the provider should close the patient’s gown and have the patient sit up on the exam bed. The provider should share any findings, e.g. masses found, with the patient while taking care not to unnecessarily upset the patient, and to address patient concerns. The provider and patient will also discuss the next steps in treatment or diagnosis, e.g. mammograms or biopsies. Both mammograms and biopsies are viewed as fearful experiences by significant portions of the population, with reasons of fear of finding cancer, pain, and radiation being significant barriers to women getting regular screening mammograms. This is especially prevalent in African American and Hispanic populations. Similarly, patients fear that if the mass is cancerous, the biopsy will spread cancer throughout the breast [103]. The provider must be prepared to address these fears when discussing these diagnostic tests. In our
evaluations of MRIPS-CBE, we have created scenarios in which the patient is fearful of mammograms, because they may be painful and because the patient associates a mammogram with causing her mother’s breast cancer, and scenarios in which the patient is fearful of a biopsy because it may spread cancer throughout the breast.

3.2.2 Cognitive, Psychomotor, and Affective Components

With respect to the skills that must be trained during the CBE, the CBE is a compound affective, cognitive, and psychomotor task – these three skill sets must be trained in a way that the healthcare provider is able to actuate all three concurrently.

The affective components of CBE are taking the patient’s perspective (perspective taking [40]) and displaying appropriate empathic behavior. These two elements allow the healthcare professional to comfort the patient, obtain an optimal amount of information disclosure and following of the healthcare professional’s treatment instructions, achieve rapport, and form a close doctor-patient relationship. Touching of the patient is often used to convey that the healthcare professional understands the patient’s perspective, as a component of empathic behavior, and as a means of comforting the patient [34][35][36][104].

The cognitive components of the CBE are split between the medical history and the physical exam. In the medical history, the cognitive components are following proper procedure – asking the correct questions, integrating the information divulged by the patient to arrive at a correct differential diagnosis and diagnostic workup (further diagnostic steps, e.g. a mammogram, and treatment plan).

In the physical exam, the cognitive and psychomotor components can be elucidated by labeling each component with a keyword from Bloom and Simpsons’ taxonomies of cognitive and psychomotor tasks [105][106]. The cognitive tasks are to
recall a procedural pattern, the pattern-of-search, in which the breast should be palpated, and the search task for unseen breast masses – recognizing which areas of the breast remain to be palpated in order to ensure complete coverage of the breast, and interpreting whether breast tissue feels like normal tissue or abnormality. The psychomotor task is palpation of the breast in circular motions with a specific motion and palpation pressure (at the low, medium, and high levels of pressure). This psychomotor task starts out as a guided response (trial and error) and must be practiced until it becomes an overt response (accurate and skillfully performed).

3.2.3 Current Approaches to Teaching Clinical Breast Examination

Accepted approaches to teaching and learning CBE follow a progression of: lectures covering technique, practice with anatomical models, e.g. a silicone breast, and practice on standardized human patients (SPs), actors trained to portray a patient with a breast complaint [9][107]. This progression takes place in the 2\textsuperscript{nd} or 1\textsuperscript{st} and 2\textsuperscript{nd} years of medical school, after which, in the beginning of the 3\textsuperscript{rd} year, students begin seeing patients in clinic. In the 3\textsuperscript{rd} year, students’ education in CBE stems entirely from experiences in clinic.

Though the individual approaches in this progression have been validated as providing an improvement in learning over lectures alone, e.g. [62][65][107][108], individual approaches have significant drawbacks. While practice with anatomical models allows learners to experience abnormal findings in an exam, e.g. a breast mass, this approach does not provide the context of a patient interaction that is critical to succeeding in examinations of human patients [7]. Students are unable to practice interpersonal skills or the merging of interpersonal and physical exam skills in this approach. Additionally, feedback on performance is rare, and typically consists of
verbal feedback provided by peers [62]. Practice with SPs allows students to focus on both interpersonal and physical exam skills, however, availability of SPs for intimate exams is scarce and SPs with abnormal findings, e.g. a breast mass, are unavailable. Opportunities for practice on SPs are further limited because of high monetary cost. Because students get few opportunities to practice with SPs, the merging of the interpersonal and cognitive and psychomotor physical exam skill sets may not occur through practice with SPs – indeed, it is common that students performing intimate exams on patients in clinic have low confidence in their physical exam skills and anxiety pertaining to both the intimate, interpersonal nature of the exam and the performance of the physical exam itself [4][38]. Additionally, novice students report fear of hurting the patient and fear of missing a lesion as causing significant anxiety during these exams, enough to prevent them from performing an effective exam and communicating with the patient. The fear of missing a lesion persists even in graduating medical students [38].

In the 3rd year of medical school, as students treat patients in clinic, their education in intimate exams comes primarily or completely from their experiences in clinic. Evaluation of students’ intimate exam skills in the 3rd year is not standardized, but a common and accepted approach is to receive feedback from an expert observer of the student’s exam [9]. This feedback has been shown to improve students' physical exam skills in future intimate exams [61][62]. However, the requirement of an expert clinician or educator to provide feedback reduces the frequency of these evaluations. Medical educators have identified a need for increased number of practice opportunities and the incorporation of standardized, immediate, objective, and detailed feedback [9]. MRIPS-CBE is targeted to train medical students prior to entering clinic in the 3rd year, to
provide novice students with additional practice opportunities and feedback of their quantitatively measured performance, with the goal of improving cognitive, psychomotor, and affective components of their CBE performance.

3.3 MRIPS-CBE

MRIPS-CBE applies the MRIPS design principles to simulation of the interpersonal scenario of clinical breast examination. As an intimate exam, CBE is a high-stakes interpersonal scenario in which it is not acceptable to fail. Using MRIPS-CBE, novice learners of CBE can practice cognitive, psychomotor, and affective CBE tasks in a learning environment in which it is safe to fail.

3.3.1 Motivations and Goals

Together with mammography and breast self examination, the clinical breast exam (CBE) is an essential part of screening for the early detection of breast cancer [9][44]. Each year in the United States, 200,000 women are diagnosed with breast cancer and 40,000 die of the disease. The CBE is critical in early detection of the disease as CBE is effective at younger ages than yearly mammography is recommended. Additionally, CBE can find masses not found by mammography [44] – each year, up to 10,000 otherwise undetected cancers can be detected through CBE – and it is not cost effective to make mammography the first step in screening [4]. Because of these factors, competence in CBE is required of all health care professionals – both in the physical exam and in communication skills to assess patient risk, reduce patient anxieties, and address findings of the exam.

Clinical breast exam is thus a high-stakes interpersonal scenario. The prevalent means of learning CBE is “live-fire” practice on real patients in clinic – this is not a learning environment in which it is safe to fail. This high-stakes aspect along with the
intimate nature of the scenario and complex set of cognitive, psychomotor, and affective skills required for success results in medical students and even practicing clinicians expressing low confidence in their CBE technique and fear that they will miss a breast abnormality [9][38]. Teaching and learning of CBE is hampered by a high degree of learner anxiety [38], a lack of standardized, objective, precise feedback, and limited opportunities for practice before the “live-fire” real patient interactions [9].

The goal of MRIPS-CBE is to provide novice learners of CBE with additional opportunities to practice the cognitive, psychomotor, and affective components of CBE and to receive objective, precise feedback concerning these three components (see Chapters 6 and 8 for a description of this feedback). The coupling of additional practice with feedback and reflection is targeted to improve learner performance in CBE of human patients (evaluated in Chapter 9), potentially resulting in more effective early screening for breast disease.

We next describe the design of MRIPS-CBE using the merging and augmenting paradigm presented in section 3.1, describe the affordances of MRIPS-CBE with respect to the cognitive, psychomotor, and affective components of CBE, and then present the implementation of the MRIPS-CBE visual interface, haptic interface, and mixed reality human simulation.

3.3.2 Merging and Augmenting in MRIPS-CBE

The physical interface of MRIPS-CBE takes the form of a mannequin with a silicone breast in the place of the left breast, and wearing a hospital gown. The mannequin’s right arm incorporates servo motors allowing it to move across the torso and touch the user. The physical interface is registered to the pose of the life-sized virtual human lying on her back on a virtual hospital bed. This affords touching of the
virtual human when she is lying on the bed – the pose in which the exam takes place and in which communicative touches are most likely (e.g. to instruct to put her arm behind her head to make it easier to identify breast tissue and palpate the armpit; to comfort when the patient expresses fear over the exam or anxiety about having her breast touched).

The appearance of the virtual human is augmented by a video stream that shows the user's hands, the silicone breast of the physical interface, and the hospital gown; the rest of the physical interface is segmented out of this video stream.

The physical interface is augmented with 64 force sensors which report touching of the virtual human. Manipulation of the hospital gown is tracked using infrared optical tracking, reporting when the gown is opened or closed on the left side (i.e. whether the left breast is exposed).

The mixed reality human simulation captures the user's speech, head pose, force sensor data, and gown manipulation data. This simulation drives outputs: virtual human speech, facial expressions, and gestures, and movement of the mannequin's right arm.

3.3.3 Cognitive, Psychomotor, and Affective Affordances of MRIPS-CBE

MRIPS-CBE attempts to afford performance of all of the cognitive, psychomotor, and affective components of CBE.

MRIPS-CBE affords the cognitive task of conducting a medical interview (assessing the virtual human patient’s medical history and risk factors for breast disease) through user speech and virtual human speech and nonverbal responses.

The psychomotor task of palpating the breast in three levels of pressure is afforded by the inclusion of a silicone breast into the physical interface of MRIPS-CBE. This breast mimics the feel and soft-tissue deformation of a human breast (though it
does not mimic thermal qualities), allowing the user to perform the psychomotor palpation task as he would on a human patient. Feedback of the quality of this psychomotor task is also presented using the visual interface (feedback described in Chapters 6-7).

The joint cognitive-psychomotor task of recognizing when the entire breast is palpated and maintaining a specific pattern of search can be performed by feeling the breast in the physical interface and viewing the breast in the visual interface. Feedback is also integrated into the visual interface to aid these tasks (see Chapters 6-7).

The affective task of comforting the patient can be engaged in using MRIPS-CBE. The mixed reality human simulation incorporates specific conversational topics and touch-triggered scenarios (e.g. opening the patient’s gown to begin palpation) which cause the virtual human to prompt (verbally and nonverbally; but never explicitly) for comfort. To comfort the patient, the user can verbally express understanding of her situation or touch the shoulder or upper arm of the physical interface (comforting touch). In response to the user’s attempt at comforting the virtual human, the virtual human expresses that she is (or is not) comforted with speech or through visually-expressed thoughts (“thought bubble” feedback discussed in Chapter 7).

**3.4 Visual Interface**

The visual interface consists of a life-sized virtual human character as well as elements of the real world that (a) augment this VH character (e.g. physical simulators, physical clothes and other objects) and (b) provide the user with high-fidelity self-avatar in the virtual world.
3.4.1 Life-sized Virtual Human Characters

The life-sized virtual human character has facial animation and gestural capabilities including lip-synched speech, eye blinking, breathing, pointing, idle behaviors (e.g. swinging legs, looking around), the ability to maintain eye contact with the user, as well as scenario specific keyframe-based animations. The virtual character is able to gesture through offline-created keyframe-based animations. The virtual character’s facial expressions consist of real-time morphing between offline-created meshes. The virtual character’s appearance and animation capabilities were designed using Autodesk Maya and Di-O-Matic Facial Studio by other researchers at UF CISE.

Specific to MRIPS-CBE, the virtual character has keyframe-based animations specific to the scenario of a breast exam. She is able to point to locations in her left breast where she has found a mass and where she is experiencing pain. She also has animations to transition between four positions used in a breast exam: sitting with arms relaxed, sitting with hands pressed on hips and chest flexed, sitting with arms raised above head, and lying down with arms raised above her head. Her facial expressions include neutral, happy (smiling), sad (frowning), confused (this is occasionally used when user speech is not recognized), in-pain (grimacing), and fearful (a combination of sad and in-pain).

3.4.2 Augmenting the Virtual World with Real Objects

Through additional sensors (webcams) and computer-vision techniques, elements of the real world are made to augment the virtual human, and the user is given a visually faithful self-avatar. Two webcams, one seeing in color, and one seeing in infrared, are mounted above, looking down upon, the physical interface. For a general application of MRIPS, the color webcam (640x480 resolution, 30Hz) is used to
incorporate a visually faithful self-avatar of the user consisting of real-time video of the
user’s hands and forearms. Other augmentations of the virtual world and the virtual
human are application specific. In MRIPS-CBE, the same color-seeing webcam and
computer vision techniques are used to augment the virtual human with a physical
breast and a physical hospital gown.

Independent of application, we are motivated to allow the user to see his hands
when touching the physical interface, as this feature was demanded by pilot study
participants [82], and it has been shown that seeing one’s hands in a virtual world
improves the believability of the experience [109]. Additionally, it has been found that
when conducting hands-on tasks in a virtual environment, users prefer a visually faithful
avatar to a generic avatar – i.e. a user wants to see his hand, not a generic hand [110].

However, providing the user with a visually faithful avatar in a virtual world is
difficult to do in an unconstrained scenario – the user’s hands must be segmented out
from the real world and incorporated into the virtual world in the position and orientation
of the user’s hands – the resulting avatar is often noisy, containing real-world pixels
which do not belong to the user’s body [110]. This is not acceptable in many scenarios
we anticipate applying MRIPS – e.g. in the scenario of a medical physical exam, the
user must see how his hands touching the patient deforms the patient’s tissue; this
deformation can not be observed if real-world background pixels occlude the virtual
patient’s tissue.

Fortunately, in our scenario of performing a clinical exam, the user only needs to
see his hands when touching the VH. This constrains the problem to providing the user
with a visually faithful avatar when his hands are within a small volume surrounding the
MRIPS physical interface. We take advantage of this constraint to simplify the segmentation task. All objects in the area of the physical interface, and all parts of the physical interface, which are visible in the webcam images used for augmentation and which should not be visible in the virtual world, are colored black. This allows these objects to be segmented from the background with little noise.

The segmentation proceeds as:

1. Perform a Gaussian blur (3x3) on the incoming camera image to remove noise within areas that should be marked as foreground.

2. Perform color segmentation, where background pixels are defined as those having luminance less than some predefined threshold (fixed at the time the system is setup, depending on real-world lighting conditions). For each background pixel, set its alpha channel value to zero. Alpha blending will be used to incorporate the final image into the virtual world – areas of the image with an alpha value of zero will not be visible.

3. Use a 3x3 mean filter to blur the alpha channel, reducing noise at foreground edges.

One benefit of this approach allows desirable real-world shadows (providing important depth cues) to be included in the virtual world – and undesirable real-world shadows (over the black-colored areas) to be discarded. In my experience, this would be difficult using color-based or background-subtraction-based image segmentation in an unprepared environment. Furthermore, because the background color is known a priori, the range of colors which the algorithm relegates to the background is very small and is known to be disjoint from the range of human skin colors. This allows the algorithm to succeed for any user, no matter how dark skinned.

Once the image of the desired augmentation is extracted from the current video stream image, the augmentations must be incorporated into the virtual world. Because the pose of the camera which performs this augmentation is fixed (and is not the same pose as the user’s head pose), we take a simple approach to displaying real
augmentations of the virtual human character in the virtual world at a believable approximation of the user’s perspective. This approach is to project the processed video stream, which provides these augmentations, onto a polygonal mesh of the physical interface. This mesh is obtained through the 3D-reconstruction technique of laser scanning the physical interface elements (breast, mannequin). The laser-scanned mesh is registered in the virtual world with the virtual character, and the video stream is a projected texture, projected from a virtual camera having a similar pose as the physical webcams, onto this mesh. This method provides depth and occlusion cues in the virtual world, without requiring video to be taken from the user’s viewpoint.

Specific to MRIPS-CBE, augmentation of the virtual world other than the user’s self-avatar consists of augmenting the virtual human character with a physical breast model and a physical hospital gown. The advantage of augmenting the virtual human character with a physical breast model is that users can be provided with realistic appearance and behavior (realistic soft-tissue deformation) of the virtual breast. In a user study we conducted to evaluate the importance of the visual fidelity of the virtual human [101], it was found that providing a realistic virtual breast resulted in increased believability, realism, and perceived educational benefit of the simulated interaction.

3.4.3 Display of the Visual Interface

The visual interface is displayed using one of two display devices: a head-mounted display or a projection screen (Figures 3-1 and 3-3). The goal of each of these displays is to allow the user and the visual interface to occupy the same physical space, allowing the user to see his touching of the virtual human in-situ with the physical silicone breast being touched. The head-mounted display (HMD) embeds the user
within the virtual world, while the projection display provides a “window into the virtual world”, embedding the virtual human within the real world [14].

3.4.3.1 Head-Mounted Display (HMD)

The HMD used is an eMagin z800 with 800x600 resolution and 40-degree diagonal field of view. This HMD is light weight, <0.5 lbs, and has an unobtrusive form-factor, making it less encumbering to wear than higher resolution and field of view (and weight and size) HMDs. When viewing the MRIPS visual interface using an HMD, haptics and visuals are co-located. The MRIPS visual interface is co-located with the MRIPS physical interface – the visual and physical interfaces occupy the same volume in 3D-space. Thus the HMD display device provides an experience most similar to the real world, at the expense of encumbering the user (covering the user’s face, wires that attach the HMD to a PC restricting the user’s movement).

3.4.3.2 Projection display

The MRH is projected on a planar surface – causing the MRIPS visual component to appear non-co-located with its physical component, but instead in a configuration similar to that of Figure 3-3. Using the projection screen display, users touch the virtual human in one volume and see the virtual human in a disjoint volume – haptics and visuals are not co-located in the real world, but the virtual human and haptic interface are still conceptually co-located (to touch the virtual human, one touches the haptic interface). The projection screen display provides an experience dissimilar from the real world, but with the advantage of unencumbered movement of the user. We have conducted a user study which compares the HMD and projection displays on the dimensions of usability, acceptability, and user behavior; no significant differences were
found, demonstrating that despite their differences, the displays are equivalently appropriate for displaying the MRIPS-CBE visual interface.

3.4.4 Perspective Correct Viewing of the Visual Interface

Independent of which display type is used, the user’s head pose (6 degrees-of-freedom: 3 degrees-of-freedom position, 3 degrees-of-freedom orientation) must be tracked to provide the user with a perspective-correct view of the virtual world, and for the user’s head pose to be an input into the simulation (e.g. allowing the virtual human to make eye contact with the user). When using the HMD, one infrared-reflective tracking marker is affixed to the HMD. Tracking of this marker by an infrared tracking system [14] composed of two infrared-seeing NaturalPoint Optitrack cameras provides the user’s head position. The user’s head orientation is provided by an inertial sensor, an Intersense InertiaCube 2, mounted on the HMD. When using the projection screen display, the user wears a hat, to which three infrared-reflective tracking markers are attached. These markers are tracked to calculate the user’s head position and orientation, which allows the virtual human to be rendered from the user’s perspective.

3.4.5 Registering Visual and Physical Interfaces

The physical interface is registered to the visual interface in a one-time calibration step. It is this calibration that allows the user to see the virtual human character in-situ with its physical representation. At setup time, an infrared marker is placed on the webcam overlooking the physical interface, and an additional infrared marker is placed on the physical interface at a location that corresponds to a known 3D-coordinate in the virtual world (e.g. for the breast exam MRH, this marker is placed on the nipple, the position of which is known in the virtual world). By tracking these two infrared markers using the same infrared tracking system used to track the user’s head pose, the position
of the physical interface and the webcam in the virtual world are calculated. The error of the registration of the visual interface to the physical interface is estimated at ~1cm at the nipple. The error is higher for other parts of the physical interface, as the shape of the physical interface and virtual human do not match in all aspects.

3.5 Physical Interface

The physical interface represents a portion, or all, of the virtual human’s body. The physical interface provides a focal point for touch of the virtual human for communication and psychomotor task performance.

In MRIPS-CBE, this physical interface takes the form of a semi-articulated (neck, arm, and elbow joints) torso-arms-and-head mannequin having a human form. The mannequin’s left breast is a silicone breast with the feel of human breast tissue. The right arm of the mannequin is replaced with a mechanical arm capable of simple movements to touch the user. The mannequin is covered with a “skin” that detects the user’s touch.

3.5.1 Active Sensing of User Touch

The skin of the physical interface consists of two thin layers (~3mm thickness) of high density foam. Between the two foam layers is a layer of force sensing resistors. The initial prototype contained only 10 force sensors and provided limited sensing area and resolution. The latest version of the physical interface of MRIPS-CBE contains 64 force sensing resistors (Figure 3-4), affording dense sensing in the full area required for a clinical breast exam and less dense sensing in areas that may be touched for communication, such as the upper arm and shoulder. The resolution at which touch is sensed in the breast is approximately 2cm, though an interpolation technique was
developed which provides continuous sensing of touch throughout the breast (Chapter 7).

The force sensing resistors are placed in circuits which cause each force sensor to output a voltage (0-5 volts) that varies (approximately) linearly with the force exerted on the sensor. The outputs of these circuits are sampled by 8-bit analog-to-digital converters at 60 Hz. This process is controlled by a Motorola 68HC12 microcontroller, which transfers the digital output of the 64 force sensors to the simulation module over a serial link. Through this process, the user's touch of the physical interface becomes an input to the mixed reality human simulation.

Before the sensor values are received by the mixed reality human simulation module, a baseline for each sensor is found as the average of its values over a period of 30 seconds (in which the physical interface is not touched). This baseline may fluctuate during an exam due to noise or a sensor becoming “stuck” (after being compressed in by a palpation, thicker areas of the silicone breast may take a few seconds of not being touched in order to become uncompressed). “Stuck” sensors can negatively influence detection of user touch and measurement of user psychomotor task performance. To address this, the value of each sensor is input into a noise gate, which produces an output value of zero if the sensor is “stuck” above its baseline for a period of time, but allows the sensor to contribute its values once it becomes “unstuck,” indicated by the value dropping below the baseline for a similar period of time.

User touch of the physical input is also detected using computer vision techniques. The infrared-seeing webcam mounted above the physical interface tracks infrared markers which are affixed to physical objects which augment the physical interface (e.g.
the physical hospital gown worn by the breast exam patient). By tracking the position and other aspects of these physical objects, the manipulation of these physical objects become inputs to the simulation (e.g. opening and closing of the physical hospital gown worn by the breast exam patient).

3.5.2 Passive Detection of User Touch and Manipulation of Tools and Props

User touch of the physical interface, which is not a direct touch of the virtual human's body (e.g. touching the virtual human's clothes, manipulating other real objects such as a stethoscope) is tracked using the infrared-seeing webcam mounted above the physical interface. Components of the physical interface, such as the virtual human's clothes and tools used to interact with the virtual human, are affixed with infrared-reflective tracking markers. In MRIPS-CBE, this approach is used to track a hospital gown worn by the virtual human.

The virtual human breast exam patient wears a hospital gown which opens in the front. This is accomplished by having a physical hospital gown be worn by the physical interface's mannequin, and a virtual correlate (textured mesh) to this physical gown be worn by the virtual character. The gown is an integral part of the breast exam: both breasts must be visually and physically examined, but to maintain patient comfort, only one breast should be exposed at a time. The gown provides the user with haptic feedback, and the opening and closing of the gown are momentous events in the interpersonal scenario of the clinical breast exam (e.g. a patient fearful of having the exam performed may exclaim “Wait, I'm scared!” when the user is detected opening the gown).

We explored two approaches to tracking the physical hospital gown in order to render a visually faithful virtual counterpart. The first approach was to use background
subtraction to create an image which indicated the area of the webcam image occupied by the gown (Figure 3-5, A). The gown is segmented from the webcam image using a Gaussian-model background subtraction approach (see [111] for a review of background subtraction techniques). This produces a binary foreground image. To reduce noise in the segmentation caused by shadows cast by the user, the binary foreground image is passed through a variation of a smoothing filter. This filter classifies an image region as foreground if the region contains a number of foreground pixels greater than a predefined threshold, and classifies the region as background otherwise (Figure 3-5, B). The resulting binary foreground image is ANDed with the visually faithful gown texture. This texture is applied at 30Hz (maximum webcam frame rate) on top of the virtual character’s mesh using multi-texturing. This allows current configuration of the physical gown to be displayed as a virtual gown covering the patient (Figure 3-5, C). Using this method, it could be determined whether the gown was open or closed based on the number of background pixels in the binary image.

Although this method provided both a virtual correlate to the physical gown, and also detected gown opening and closing using a single color webcam, the resulting visuals are less realistic and convincing than taking the approach used to produce the visuals in Figures 3-1 and 3-3. These visuals incorporate the second approach to tracking the gown – the color webcam uses color segmentation to make the physical gown augment the virtual world (similarly to providing the user with a self-avatar), and an infrared webcam, with a similar viewpoint as the color webcam, tracks a strip of infrared-reflective tape placed on the physical gown to determine if the gown is open or closed.
The edges of the gown (the infrared-reflective tape) are segmented from the image by color segmentation (the IR reflective edges show up as white in the image, while the background is grey/black). Before the user begins the examination of the patient, a line is fit to the foreground pixels, providing a baseline of where the edges of the gown are located when the gown is closed. For subsequent frames, as the user manipulates (opens, closes) the gown, a line is fit to the foreground pixels using least squares. The line of the current frame and the baseline frame are clipped to the edges of the image. This results in four points which form a quadrilateral which is an estimation of the open area of the gown. The area of this rectangle is calculated and compared to a pre-set value in order to determine if the gown is open or closed. The advantage of this method is that it produces a higher fidelity visual representation of the virtual gown while detecting opening and closing with less noise than the first method described.

3.5.3 Bidirectional Touch: Enabling the Virtual Human to Touch the User

We have taken two approaches to allow the VH to make an interpersonal touch of the user for the purpose of communicating: purely virtual touch and physical (active-haptic) touch.

3.5.3.1 Purely virtual touch

The first approach is to incorporate the virtual human’s touching of the user into the visual interface: the virtual human character is given an animation in which the virtual human’s hand appears to touch the user’s self-avatar. This approach is known as visual touch or pseudo-haptics [98][99]. In this approach, the perception of a haptic stimulus from the purely visual stimuli is an effect of synesthesia, a physiological
phenomenon in which data from one sense (e.g. vision) fills in missing data of other senses (e.g. touch).

3.5.3.2 Physical touch

The second approach to allow the virtual human to make a communicative touch of the user is to provide the virtual human a means of physically touching the user. This active-haptic touch is provided by a simple mechanical arm which augments the physical interface. This arm was designed for the specific purpose of the MRIPS-CBE patient touching the doctor on the hand during the clinical breast exam (Figure 3-6).

The arm consists of upper arm and forearm-hand sections connected by a servo motor. Another servo motor attaches the shoulder of the arm to the instrumented mannequin of the tangible interface. The two servo motors provide 180 degrees of flexion/extension motion at the shoulder and elbow joints. A passive joint at the wrist and padding of the hand reduce the force of the moving arm to that of a social touch (Figure 3-7).

The mechanical arm uses two Hitec HS-5645 MG digital servos which are capable of 168 oz.* in. of torque and 60° of rotation in 0.18 seconds. Desired rotations of the digital servos are achieved by having a hardware servo controller send a pulse-width modulated signal to the servo [112]. The duty cycle of this signal specifies the absolute angle the servo should be rotated to. The controller and servos are abstracted by a high-level framework which allows movement of the virtual human’s mechanical arm to appear to the controlling application to be a keyframe-based animation, similar to those used to animate the virtual human. These “mechanical animations” are defined off-line as a series of sequential or concurrent rotations of servos. At run-time, the controlling
application calls the desired animation, resulting in the predefined movement of the arm to physically touch the user.

To enable the movement of the mechanical arm to appear to the application as a simple animation call, an extensible framework was designed. This framework is a high-level API allowing a virtual reality application to control locally or remotely located groups of actuators, including servos, as a means for incorporating mechanical devices into virtual experiences. The framework is visualized in Figure 3-8.

At the application layer, the movement of the mechanical devices is abstracted as an animation, allowing control of mechanical devices in virtual reality applications to be conceptualized in the same manner as traditional animations in computer graphics (e.g. runAnimation("point to user")).

The actuator abstraction and actuator network abstraction layers abstract a single actuator and a network of actuators, respectively. An actuator is an electro-mechanical device (such as a servo motor) which takes as input an electrical signal and produces a mechanical output (physical motion). An actuator network consists of one or more locally or remotely located actuators. Communication between remote systems is enabled through the Virtual Reality Peripheral Network library VRPN [113]. At this level, the animation is decomposed into individual commands given to actuators described by a template of <action, actuator, value> (e.g. "rotate servo_elbow 90°"). This level also enforces constraints of the networks of actuators. Constraints allow for simulation of skeletons of actuators. A skeleton defines relationships between actuators in a network and the conditions these relationships impose. For example, the mechanical arm of the MRH breast exam patient has a skeleton consisting of two servos:
servo_shoulder_extend and servo_elbow_extend. The relationship between these two actuators is that servo_elbow_extend should only operate if servo_shoulder_extend has an angle of greater than 114° (highlighted in yellow in Figure 3-12).

The hardware controller abstraction is a software interface written for a generic hardware controller. It is at this level that the abstraction of having an actuator perform some action is translated into, e.g., having a servo perform a rotation.

The hardware controller interface is a software interface written for a specific hardware controller or set of controllers. This layer incorporates any APIs of commercial controllers used. For example, this layer translates rotateServo(String servoname, float degree) into a hardware specific API call phidgets_api_move_servo(int controllerId, int servoid, float degree), where phidgets is a set of commercial hardware controllers.

These layers fit within a mechanical device controller application, which automates initialization of all hardware controllers used in the active animation, and runs all commands of the active animation, specifically handling thread management allowing for concurrent control of multiple actuators.

As part of the evaluation of the MRIPS-CBE and its usability for training communication (cognitive and affective) aspects of CBE, we conducted a user study to determine if pseudo-haptic (purely virtual) or active-haptic touch (physical touch) were preferred for allowing the virtual human to initiate interpersonal touch and communicate with the user. Participants indicated that both approaches were successful at allowing the virtual human to communicate, but that active-haptic touch may be more effective for communication and was rated as more realistic [93].
3.6 Mixed Reality Human Simulation

The mixed reality human simulation module is an extended version of the simulation module used in the interpersonal simulator of Johnsen et al. [10].

By taking in the inputs and producing the outputs listed in Table 3-1, the simulation module affords bidirectional verbal, gestural, and haptic communication between virtual human and human. The design of the simulation module is shown in Figure 3-9. For each scenario, an xml database of triggers and responses is created by a scenario designer with input from domain experts. The function of the simulation is to map simulation inputs to a trigger, which, as defined in the xml database, is in turn mapped to a set of responses. These responses are executed by the rendering module, resulting in the simulation output – the speech and nonverbal behavior of the virtual human character of the visual interface.

To map the simulation inputs to triggers in the xml database, the current set of inputs and the current simulation state are examined. Each trigger can be thought of as a vector of trigger variables $T$, and a vector of pre-condition expressions $P$ (e.g. “user_said_hello != 0”). For each trigger, if $T$ is found to be a subset of the simulation inputs $I$ and $P$ is found to be a subset of the current simulation state $S$, the trigger is activated and the corresponding responses $responses(T)$ are executed.

At each timestep in the simulation, the function of the simulation module is that of the following pseudocode method $match(\text{Triggers})$.

Responses $R \leftarrow NULL$
For each pair $(T,P)$ in Triggers,
    For each pre-condition $p$ in $P$,
        If $p \notin S$, exit
        Else, continue
    For each trigger variable $t$ in $T$,
        If $matchSubset(t, I)$, continue

103
Else, exit
\[ R \leftarrow responses(T) \]
\( \text{execute}(R) \)

Each type of input listed has a different matching algorithm \((\text{matchSubset})\) associated with it. Matching of speech and user tracking data has previously been described by Johnsen [14]. The speech matching algorithm matches text of the user's recognized speech to keywords defined in the trigger variable vector \(T\).

My contributions to the simulation module have been to add a flexible system of maintaining state, and to incorporate haptic interaction inputs (user touch of the physical interface and tracking of the manipulation of physical interface items) into the simulation.

### 3.6.1 Maintaining and Applying Simulation State

Simulation state is maintained in a simple fashion – each set of executed responses \(R\) contains a vector \(S_R\) containing dyads \(<\text{function } f, \text{ value } v>\). After executing the response, the new value of each state variable \(s\) in the system state vector \(S\) is updated as, \(s \leftarrow f(v)\). By allowing each trigger to require that some subset of state variables have specified values (i.e. have pre-conditions), the interaction between the human and VH can be directed down certain paths, or be allowed to proceed in a free-form manner.

Providing the simulation with the ability to direct the human-virtual human communication is necessary when the space of inputs is expanded from only speech input to combinations of speech input and haptic input (e.g. touching the virtual human while asking “does this hurt?”).

In the example shown in Table 3-2, each of the user's statements is extremely similar – so similar that the keyword matching system will likely not be able to tell them
apart – and these statements should elicit different responses from the virtual human. The virtual human should not provide the 2\textsuperscript{nd} response if she has not already provided the 1\textsuperscript{st} response. The current state of the simulation after the first virtual human response is taken into account when the 2\textsuperscript{nd} user speech is processed, and results in the appropriate 2\textsuperscript{nd} virtual human speech. This example illustrates the ability of maintaining simulation state to provide the virtual human with the appearance of understanding the context of the user’s speech. The simulation state is designed to be flexible in its use, e.g. it can also be used to keep the virtual human from performing a certain action until the user has performed prerequisites for this action. We use this function of the simulation state to guide the user and virtual human through a three part exchange when the user recommends the virtual human receive a mammogram. The virtual human first expresses fear that the mammogram will hurt, then expresses fear that it will find cancer, and, with further prompting from the user, agrees to have a mammogram performed. Keeping track of simulation state can also allow for asynchronous arrival of speech and touch inputs which are part of the same query to trigger the correct response (e.g. touch followed 2 seconds later by “does this hurt?”; “does this hurt?” followed 3 seconds later by touch; these can both trigger the correct response: “it’s a little tender there”, even though the speech never arrives at the exact time as the painful area is being touched.

3.6.2 Incorporating User Touch of the Physical Interface into the Simulation

Haptic inputs to the simulation are the user’s touch of the physical interface, and the user’s manipulation of other physical objects. User manipulation of physical objects is handled by the simulation just as any other position-orientation tracking data input to the simulation. User touch of the physical interface is handled through simple “touch
templates” which define the mapping of low-level sensor inputs to high-level “touch gestures.” E.g., the low-level sensor inputs “sensor4=2.0v; sensor5=2.5v” are mapped to the gesture “user_touching_mrh_patient’s_left_breast” by this template:

```
<touch-template name="user_touching_mrh_patient’s_left_breast">
  <sensor name="sensor4" value="1" operation=">=" link="and"/>
  <sensor name="sensor5" value="2" operation=">=" link="or"/>
</touch-template>
```

Using a simple hierarchical approach, the simulation module matches the current set of sensor values to a touch-template. From a software engineering standpoint, it is valuable to take a hierarchical, multi-level approach, in order to provide extensibility of the implementation. However in practice, we have found that a single level (flat hierarchy) is sufficient for most touch gestures. Only a handful of higher-level gestures benefit from the multi-level approach (certainly these can also be described in a single level, but creation of the xml script becomes unwieldy), e.g.:

```
<touch-template name="touch_exam_related">
  <sensor name="touch_breast" value="1" operation="==" link="and"/>
  <sensor name="touch_neck" value="1" operation="==" link="or"/>
</touch-template>
```

Above, touch_breast is the template described in Listing 3-3, and touch_neck is another template not shown that involves many sensor inputs. The touch-gestures which are the output of this matching step are trigger variables which are matched to triggers using the matching algorithm for which pseudocode is provided in `match(Triggers)`.

### 3.6.3 Touch-Driven Communication

The design of MRIPS affords bidirectional verbal and nonverbal communication between human and virtual human. By providing this verbal and nonverbal communication including communication through touch, MRIPS provides a solution to
the difficult human-computer interaction problem of simulating a social interaction between two humans.

- Communication through touch is afforded by a feedback loop. This feedback loop proceeds as:
  The user touches the VH.

- The resulting values of the touch sensors are examined to recognize the corresponding “touch-gesture.”

- The corresponding touch-gesture is matched to a VH response.

- The VH communicates with the user as a result of the user touching the VH (the VH may also visually or physically touch the user in response).

- The VH’s verbal, gestural, or haptic response leads the user to his next verbal or haptic communication.

- The following example of touch-driven communication was incorporated into one of the user studies described in Chapter 4. When the user begins to palpate the patient’s breast, the user’s touch is detected by the force sensing resistors (for illustrative purposes, the touch is detected at sensors 4 and 5); the sensor values at sensors 4 and 5 are matched to the touch-gesture “user_touching_mrhpatient’s_left_breast”; this touch-gesture is mapped to a trigger which has as its responses the VH assuming a fearful facial expression and exclaiming “Wait! I’m scared that you might find something wrong with me… will it be ok?” The virtual human’s verbal and gestural responses should lead the user to express empathy towards the virtual human with a comforting verbal response and possibly a comforting interpersonal touch of the virtual human.

  Additional touch-driven communication events have been built into MRIPS-CBE.

During the exam, the virtual human patient is able to tell the user when a painful area of her breast is palpated (“Yeah, that hurts a little” accompanied by a grimace) and describe relative levels of discomfort between two areas of her breast being palpated (e.g. participant: “Does it hurt more or less when I press over here?” could elicit virtual human responses such as “About the same” or “It hurts the most below the nipple”).

After palpation is completed, the virtual human asks the user “does everything feel ok?” challenging the user to gently broach the subject of any masses which had been found.
During palpation, the virtual human tended to stare at the ceiling of the exam room; however, by touching the physical interface on the upper arm or shoulder, the participant could get the virtual human’s attention in the form of eye contact from the virtual human.

### 3.7 Example MRIPS-CBE Interaction

For two of the user studies conducted with MRIPS-CBE, the MRH patient was a 34-year-old female who has been experiencing a persistent pain in her lower left breast for the past three weeks. She has lost her mother to breast cancer within the past two years, and is fearful that the exam will find an indication that she too has cancer.

The user spends the first five minutes conversing with the VH patient to take a breast history (Figure 3-10). The goals of this conversation (a cognitive task) are to elicit: the patient’s current complaint, past medical history and social history, as well as her family medical history. This involves the user first introducing himself, and asking the patient’s name and age and eliciting information concerning the patient’s complaint:

User: “Hi I’m Matthew a medical student, how are you doing today?”
VH: “I’m ok, I’m a little nervous though.”
User: “Ok, may I ask your name”
VH: “My name is Amanda Jones”
User: “How old are you Miss Jones?”
VH: “I’m 34”
User: “What brings you in today?”
VH: “I’ve had pain in my left breast for a while.”
User: “How long is a while?”
VH: “I guess I’ve had pain for about a month. I’ve just been scared to come in.”
User: “Can you describe the pain?”
VH: “It’s kind of a dull pain. It’s really sensitive to the touch though.”

The user then investigates the patient’s past medical history (e.g. “Have you ever had breast pain before?”), the patient’s social history (e.g. “Are you sexually active?”), and past family history:
User: “Has anyone in your family ever had breast cancer?”
VH: “I lost my mother to breast cancer two years ago.”
User: “I’m sorry to hear that. It must make this very hard for you.”

The user then conducts a visual inspection of the patient’s breasts, asking her to first remove her gown, and then to pose in two positions: with arms raised over her head (Figure 3-11) and with hands pressed to her hips and chest flexed.

After the visual inspection, the patient is asked to lie down and put her left arm behind her head (as in Figures 3-1, 3-3). The user then proceeds with palpation of the patient’s left breast, after first opening the patient’s gown to expose the left breast (psychomotor and cognitive-psychomotor tasks are performed during palpation). When the user is first detected palpating the patient’s breast, the patient expresses fear that the examination might find something bad (cancer): “wait! I’m kind of scared about this. Is it going to be ok?” This expression of fear and prompting for reassurance should be responded to empathetically by the participant (an affective task). When finished with palpation, the user discusses his findings with the patient, stating that he found what may be a breast mass, and recommending the patient receive a mammogram for further diagnosis. The patient expresses a fear of mammograms: “aren’t mammograms painful?” and “do I really have to get a mammogram? I mean, my mom was fine, then she had a mammogram, then all of the sudden she was really sick.” This is an opportunity for the user to reassure and comfort the patient (affective tasks), while achieving compliance with the recommended diagnostic procedure of receiving a mammogram (cognitive task). After expressing that a mammogram is not painful and is important to determine what may be wrong with the patient, the patient expresses that
she will consent to a mammogram: “if that’s what you think is best.” The user then concludes the encounter.

3.8 Pilot Study

A pilot study was conducted to determine whether MRIPS-CBE is usable to perform the cognitive, psychomotor, and affective components of the CBE. The question this study posed was: “do learners apply the skills they have previously learned in real-world interpersonal scenarios to the simulated interpersonal scenario of MRIPS-CBE?”

Observed participant behavior revealed that learners of MRIPS-CBE applied the cognitive, psychomotor, and affective skills they had previously learned through traditional educational methods. Of note, participants frequently touched the virtual human to comfort her during expressions of fear and pain.

3.8.1 Population and Procedure

Eight 2\textsuperscript{nd}-year physician assistant students at the Medical College of Georgia conducted a CBE using MRIPS-CBE. These students were inexperienced in CBE. Although they had all received lecture-based teaching of CBE procedure, only one of the students had previous experience performing CBE on a human patient. All participants had conducted medical interviews of human patients (ranging from 1 to 6 interviews with an average of 2.4).

Using MRIPS-CBE, participants began with a medical history of the patient lasting approximately 10 minutes, and then performed a physical exam lasting from 5-10 minutes.

During palpation, the virtual human expressed pain when participants touched a pre-defined (but unknown to participants) section of her breast. The patient exclaimed
“ouch!” or “that hurts!” Participants’ responses to this moment and use of comforting touches were observed to evaluate whether participants were able to apply their real-world learned affective skills to the simulated interpersonal scenario MRIPS-CBE.

Participants were not given any instruction regarding touching of the virtual human in MRIPS-CBE.

3.8.2 Observations

All participants were able to successfully perform the cognitive task of taking a medical history of the patient using MRIPS-CBE. Participants discovered important pieces of information such as: the patient had found a walnut sized mass in her left breast, the patient had a family history of breast cancer (sister), the patient was postmenopausal, and the patient was taking hormone replacement.

All participants were also able to perform the psychomotor and cognitive-psychomotor tasks involved in palpation. In this pilot study we did not evaluate the quality of this task performance; participants were inexperienced and not expected to perform these well, but all participants attempted these tasks to the best of their ability.

A majority of participants performed successfully at the affective task of comforting the patient. Seven of the participants elicited the pain response from the virtual human, indicating a need for comforting through the expression of empathy. Five participants responded with empathic statements, e.g. “I know it’s tender there. I’ll try to be more gentle.” Non empathic responses from the other two participants indicated that they at least understood the patient was in pain, e.g. “Oh, it hurts right there?” Compared to a previous study of 27 2nd-year medical students interviewing a virtual human with a breast mass using the IPS system of Johnsen et al. [15] (in which touch was not afforded), more participants exhibited empathy and appropriate responses to a prompt
for comforting using MRIPS-CBE. In the prior study with IPS, 70% of participants did not respond appropriately to a virtual human’s expression of fear; only 10% responded empathetically and 20% completely ignored the prompt for comforting. The participants using MRIPS responded to the virtual human’s need for comfort in the manner they were taught and are expected to respond to a human patient; the participants using IPS responded dissimilarly from the manner they are taught to treat human patients.

Participants also used touch for cognitive and affective tasks such as instructing and comforting the patient. Seven of the eight participants touched the virtual human for communicative purposes. Each participant used an average of 1.4 (std. dev. of 0.9) of these touches. We sought to determine if this was a similar amount of touch as used in the standardized patient interactions in which these students had learned the skill of interpersonal touch. We were able to obtain data from a user study in which 76 students examined an SP with abdominal pain (data for an intimate exam scenario was unavailable). In the SP scenario, participants used an average of 1.8 (std. dev. = 1.8) touches. A statistical test of equivalence [114] indicated that it was likely that we would find an equivalent amount of touch in MRIPS and SP scenarios if a larger population was obtained for the MRIPS scenario.

3.8.3 Discussions

Participants performed all three task sets of CBE: cognitive, psychomotor, and affective. The observations made in this pilot study presented positive preliminary evidence that MRIPS presents advantages over previous interpersonal simulations, as the haptic interface of MRIPS provided touch which was used by participants to perform cognitive and affective tasks in addition to the expected psychomotor task performance. Additionally, the improvement in affective performance over the prior IPS scenario could
indicate that the additional interaction modality of touch makes users treat the MRIPS scenario more like a real-world interpersonal scenario.

### 3.8.4 Conclusion and Further Evaluation

The pilot study established that users of MRIPS-CBE were able to perform the cognitive, psychomotor, and affective components of the CBE as they had been taught.

Chapter 4 presents two additional user studies which establish that MRIPS-CBE elicits performance indicative of the learner's performance with a standardized human patient and that performance in MRIPS-CBE differentiates between users of different skill and experience levels. These studies establish the validity of MRIPS-CBE for practicing and evaluating a learner's cognitive, psychomotor, and affective performance in CBE.

After establishing that MRIPS-CBE can be used for evaluating a learner’s CBE performance, we can get to the heart of the thesis: evaluating whether a learner’s CBE performance can be improved by use of MRIPS-CBE (Chapter 9).
Table 3-1. List of simulation inputs and outputs.

<table>
<thead>
<tr>
<th>Inputs from physical sensors</th>
<th>Outputs of the virtual human simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>User speech</td>
<td>VH speech (pre-recorded human speech)</td>
</tr>
<tr>
<td>User touch of physical interface</td>
<td>VH gestures (keyframe animation)</td>
</tr>
<tr>
<td>Tracking of physical interface</td>
<td>VH facial expressions (mesh morphing)</td>
</tr>
<tr>
<td>User tracking data (e.g. head pose)</td>
<td>VH visual touch of the user (keyframe animation)</td>
</tr>
<tr>
<td></td>
<td>VH physical touch of the user (mechanical actuation)</td>
</tr>
</tbody>
</table>

Table 3-2. An example of using system state to direct the conversation between user and MRH to take a specific path.

<table>
<thead>
<tr>
<th>State (transitions are marked): value pre-speech</th>
<th>Input (user) / Output (virtual human)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Told_need_mammogram</td>
<td>Afraid_of_mammogram</td>
<td>Accepted_mammogram</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0→1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0→1</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3-1. A learner performs a CBE in MRIPS-CBE. He touches the physical interface and sees his touch reflected in the visual interface.
Figure 3-2. System design of MRIPS-CBE. The learner’s touch of the physical interface, speech, tracked head pose, and manipulation of the hospital gown are inputs into the simulation. The outputs of the simulation are the virtual human’s speech, facial expressions, and gestures.
Figure 3-3. MRIPS-CBE visual interface presented on a projection screen.

Figure 3-4. The physical interface of MRIPS-CBE. Beneath the modular foam skin and silicone breast are 64 force sensors.
Figure 3-5. The first approach taken to tracking the physical gown and providing a corresponding virtual gown used a background subtraction approach. This approach was noisy and less realistic than the video augmentation method seen in Figures 3-1 and 3-3.
Figure 3-6. Affording bidirectional touch for communication by allowing the MRH to touch the user. A) a purely virtual pseudo-haptic touch and, B) a physical active-haptic touch from the robotic right arm of the physical interface.
Figure 3-7. The physical interface incorporates a mechanical right arm, allowing active-haptic touch from virtual human to human user.

Figure 3-8. The layers of the framework for abstracting a virtual environment’s control of physical actuators.
Figure 3-9. Progression of one time step of the simulation module. At time $t$, the simulation module takes an input vector $I(t)$, the set of inputs detected by physical sensors placed on and around the user, and simulation state vector $S(t)$, and produces an output vector $O(t+dt)$, the set of VH responses and a modified simulation state vector $S(t+dt)$.

$S(t) = \langle s_1, s_2, \ldots, s_n \rangle$

$I(t) = \langle \text{speech, tracking data} \rangle$

$S(t+dt) = \langle s_1', s_2', \ldots, s_n' \rangle$

$O(t+dt) = \langle \text{VH speech, gestures, facial expressions, touch} \rangle$

Figure 3-10. The medical interview portion of the interaction with the MRH breast exam patient. The student converses with the virtual human to gain information concerning the virtual human’s medical history and current condition.
Figure 3-11. Two of the poses required for visual inspection. The virtual human raises her arms above her head and presses on her hips to flex her chest to allow the healthcare provider to inspect her breasts for abnormal appearance (e.g. asymmetry or redness).

```
<servoxml>
  <servocontroller name="phidget4" type="phidgets" serial=28304 capacity=4>  
  <servo name="elbow_yaw" location=1 rangestart=160 rangedend=50/>  
  <servo name="shoulder_pitch" location=2 rangestart=184 rangedend=50/>  
  </servocontroller>
  <skeleton name="rightarm">  
    <servo name="shoulder_pitch" controller="phidget4" parent="null" base=155>  
    </servo>  
    <servo name="elbow_yaw" controller="phidget4" parent="phidget4:shoulder_pitch" base=156>  
      <constraint servos="phidget4:shoulder_pitch" operations="&lt;", value="114"/>  
    </servo>
  </skeleton>
  <program name="swat_hand">  
    <concurrent>  
      <rotate skeleton="rightarm" servo="shoulder_pitch" angles=-69 waitfor="" waituntil="" waitcondition=""/>  
      <rotate skeleton="rightarm" servo="elbow_yaw" angles=-52 waitfor="phidget4:shoulder_pitch" waituntil="lessthans" waitcondition=113/>  
    </concurrent>  
    <sequential>  
      <wait condition="time" value="5000">  
      </wait>  
      <concurrent>  
        <rotate skeleton="rightarm" servo="elbow_yaw" angles=52 waitfor="" waituntil="" waitcondition=""/>  
        <rotate skeleton="rightarm" servo="shoulder_pitch" angles=69 waitfor="phidget4:elbow_yaw" waituntil="greaterthan" waitcondition="130"/>  
      </concurrent>  
    </sequential>
  </program>
</servoxml>
```

Figure 3-12. The xml script that defines relationships between servos, constraints, and “animation” of servos to allow the MRH to touch the user on the hand.
CHAPTER 4
VALIDITY OF MRIPS-CBE FOR PRACTICE AND EVALUATION OF COGNITIVE, PSYCHOMOTOR, AND AFFECTIVE SKILLS

This chapter describes two user studies which evaluated the potential for MRIPS-CBE to educate the cognitive, psychomotor, and affective skill sets of its users and for educators to use MRIPS-CBE to evaluate learners’ performance.

Study MRIPS-SP compared novice learners’ CBE performance with MRIPS-CBE and with a standardized human patient (SP) using a within-subjects design. Results of this study demonstrated content validity of MRIPS-CBE for practicing CBE. The validity of MRIPS-CBE for evaluating learners’ CBE skills was additionally demonstrated for cognitive and psychomotor skill sets, but not for affective skills. The lack of evidence of validity for evaluating affective skills was due to high variability in expert ratings of participants’ affective performance.

The second study, Study MRIPSx2, further evaluated the validity for evaluating learners’ CBE skills by investigating the ability of MRIPS-CBE to distinguish between learners of different skill sets on dimensions of cognitive, psychomotor, and affective performance. Expert ratings of affective performance were again used, but a simpler ratings instrument was used. Results of this study established validity of MRIPS-CBE for evaluating learners’ cognitive, psychomotor, and affective performance.

These studies serve to demonstrate that MRIPS-CBE can be substituted in place of an SP for practice and evaluation of CBEs, and motivate further evaluation to determine what learning takes place in users of MRIPS-CBE.

Collaborators: Medical collaborators Scott Lind, Adeline Deladisma, Andy Laserno, and Angela Gucwa recruited study participants and assisted in running the studies described in this chapter. Thanks to Scott Lind, Angela Gucwa, Teresa Lord,
Hevil Shah, and other clinicians and medical educators who participated in rating videos of participant interactions. Thanks goes to Adeline Deladisma for assisting in rating participant notes for correctness of diagnosis and diagnostic workup.

**Personal contributions:** I designed all studies and performed all analysis described in this chapter.

**Relevance to thesis:** The thesis focuses on demonstrating learning and training transfer with MRIPS. But, why should we expect users of MRIPS-CBE to learn (improve in) CBE or for this learning to transfer to CBE of human patients? The studies described in this chapter lay the foundation for evaluating learning and training transfer, by demonstrating that users of MRIPS-CBE were able to apply their cognitive, psychomotor, and affective skill sets in a fashion similar to CBE of a human patient. These studies demonstrated the validity of MRIPS-CBE for practicing CBE and evaluating components of learners' CBEs.

**4.1 Introduction**

This chapter describes two user studies that were conducted to evaluate the validity of MRIPS-CBE for practicing and evaluating the cognitive, psychomotor, and affective components of learners' CBEs.

Content validity of MRIPS-CBE for practicing CBE is established if learners are able to perform similarly using MRIPS-CBE and a previously validated means for practicing CBE. Study MRIPS-SP sought to establish this validity. Novice learners performed a CBE using MRIPS-CBE and a CBE of a standardized human patient (SP). Cognitive, psychomotor, and affective performance in MRIPS-CBE was found to be statistically equivalent or non-inferior to performance with the SP. Learners were able
to use the same skills with MRIPS-CBE and the SP, establishing validity of MRIPS-CBE as an additional practice opportunity for CBE.

Additionally, learners’ cognitive and psychomotor performances with MRIPS-CBE and the SP were significantly correlated, establishing the validity of MRIPS-CBE to evaluate learners on cognitive and psychomotor components of the CBE. Affective performance in MRIPS-CBE and with the SP was not correlated, due to high variability in expert ratings of learners’ affective performances. Because of this we further investigated the validity of MRIPS-CBE for evaluating CBE skills, using another approach.

Another approach to establish a simulation’s validity in evaluating real-world skills is to show that the simulation distinguishes between users of different skill levels [42]. We took this approach in Study MRIPSx2, in which 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> year medical students as well as interns, residents, and clinicians had their cognitive, psychomotor, and affective performance evaluated in a CBE using MRIPS-CBE. Results established the validity of MRIPS-CBE for evaluating learners’ cognitive, psychomotor, and affective performance in CBE.

These studies demonstrate that users of mixed interpersonal simulation treat their mixed reality human interaction partners similarly to how they treat human interaction partners, and that an interpersonal simulator incorporating haptic interfaces elicits and can be used to evaluate users’ real-world cognitive, psychomotor, and affective skills in an interpersonal scenario. MRIPS-CBE presents a novel opportunity for learning CBE which would be a valid addition to a CBE curriculum.
4.2 Study MRIPS-SP: Comparing MRIPS-CBE to CBE of an SP

This study set out to demonstrate that MRIPS-CBE can elicit cognitive, psychomotor, and affective performance similar to the standardized human patient (SP). We chose to compare MRIPS-CBE to the SP because the SP is the gold-standard for learning and evaluating CBE performance [16]. To evaluate whether learners performed similarly with MRIPS-CBE and the SP, we conducted a user study in which novice medical students performed CBEs on each of the MRIPS-CBE mixed reality human patient and an SP.

We do not present hypotheses for this study as none were written a priori. As our goal is for MRIPS-CBE to become a tool for practicing and evaluating learners’ CBE, the analysis focuses on determining whether MRIPS-CBE is valid for practice and evaluation of CBE performance. We designed the study and conducted analysis to query the:

1. Validity of MRIPS-CBE as a practice tool for CBE. If participants’ cognitive, psychomotor, and affective performances with MRIPS-CBE are statistically equivalent or non-inferior to their performances with the SP, we establish the content validity of MRIPS-CBE for practicing CBE. Content validity for practicing CBE indicates that MRIPS-CBE can be used in place of an SP for practicing CBE, as it is known that learners perform equivalently in both MRIPS-CBE and SP. Note that we only address practice of CBE; evaluation of learning is left to Chapter 9.

2. Validity of MRIPS-CBE for evaluating medical students’ CBEs along cognitive, psychomotor, and affective dimensions. Validity of an interpersonal simulation for evaluating performance is demonstrated by showing significant correlations between learner performance in the interpersonal simulation and the gold-standard for evaluation [81]. For CBE, this gold-standard is the SP. Thus if significant correlations are found between learners’ cognitive, psychomotor, and affective performances with MRIPS-CBE and with the SP, we will accept that MRIPS-CBE is valid for evaluating learners’ CBE skills.

Establishing equivalence or non-inferiority of MRIPS-CBE for practice of CBE indicates that MRIPS-CBE can be substituted for an SP interaction. As there are
logistical problems limiting the use of SPs, the ability to validly substitute a MRIPS-CBE interaction for an SP interaction can potentially increase the quantity and frequency of learners’ practice opportunities in a curriculum for learning CBE, without a reduction in the quality of the practice experience.

4.2.1 Study Design and Procedure

Ten (n = 10) medical students in their 2nd semester of medical school conducted two CBEs, one using MRIPS-CBE and the other of an SP. A counterbalanced design was used with five participants examining first the SP and then MRIPS-CBE (Group SP-MRIPS), and five participants examining first MRIPS-CBE and then the SP (Group MRIPS-SP). None of the participants had previous experience performing CBE on a human patient or SP, although all had conducted medical interviews of SPs (an average of 2.0 SP interviews).

For this scenario, the mixed reality human patient of MRIPS-CBE was a 34-year-old female who has been experiencing a persistent pain in her lower left breast for the past three weeks. She has lost her mother to breast cancer within the past two years, and is fearful that her breast pain is due to cancer. Two simulated masses were placed in the MRIPS breast model, each a hard mass with approximately 0.5 cm radius.

The SP was a middle-aged female who has found a breast mass that “comes and goes” for the last six months. Recently the mass has increased in size. Due to SP availability, three actresses played the SP. To account for this variability, the varying of the actress was balanced between Groups SP-MRIPS and MRIPS-SP. In order to evaluate participants’ ability to find masses in MRH and SP breasts, an SP with a breast abnormality was required. As SPs with breast masses are not available, participants did not perform palpation on the SP’s breasts but instead palpated a silicone breast.
model placed on the hospital bed next to the SP. This breast model was the same used in MRIPS-CBE. The SP observed the palpation and was trained to respond to the participant’s palpation of the silicone breast, expressing pain if a tender area was palpated and acknowledging the location of the mass if asked by the participant (e.g. “is this the mass you found?”). It has previously been shown that similar integration of simulators into an SP encounter provides a similar educational experience to performing the exam directly on the SP [115]. This method has also been validated for practice of physical exams and procedures which cannot be performed directly on SPs [7]. A single 0.5 cm radius hard mass was placed in the SP silicone breast. Although the number of masses in the MRIPS-CBE and SP breasts differed, finding of masses was determined to be acceptable in judging completeness of participants’ exams. The masses in the MRIPS-CBE and SP breasts were judged by collaborating medical educators as equally difficult to palpate. Since the masses are of equal difficulty to find, an exam is complete if it finds all masses regardless of the number of masses present [4][66].

Three critical moments were integrated into the MRIPS-CBE and SP interactions. A critical moment is an instance in the scenario which prompts the learner to utilize his affective skill set, to take the patient’s perspective and express empathy or concern. While explaining her symptoms, the patient expressed her fear that she could have cancer by fearfully asking “could it be cancer?” Later in the exam, when the participant began to palpate, the patient expressed fear that the examination might have a negative outcome (i.e. cancer): “wait! I’m kind of scared about this. Is it going to be ok?” After
palpation was complete, the patient inquired about what would happen to her, asking “can you tell me what happens next?”

4.2.2 Measures

Evaluating cognitive performance consisted of evaluating the completeness of the medical history and correctness of the differential diagnosis and diagnostic workup (what tests or procedures to perform after the exam).

To evaluate the completeness of the medical history, an objective, quantitative score of completeness was obtained by reviewing logs and videos of the MRIPS-CBE and SP interactions. The cognitive performance of each interaction was scored as the number of items of Table 4-1 that the participant queried. The items in this medical history completeness checklist are taken from instruments used in curricula at the University of Florida and the Medical College of Georgia, and consist of crucial topics which should be addressed to evaluate a patient’s risk factors for breast cancer. Data for this checklist is presented in Appendix A.

The correctness of the differential diagnosis and diagnostic workup were evaluated by medical educators at the Medical College of Georgia. A correct differential diagnosis for the MRIPS-CBE and SP was a malignant or benign tumor or cyst and a correct diagnostic workup must have included a mammogram.

Psychomotor performance was evaluated as the completeness and effectiveness of the physical exam. Completeness was measured as the time spent palpating, and effectiveness was measured as the number of breast masses found. The amount of time spent palpating indicates the effort put into the exam and is positively correlated with the number of masses found \([4][65]\). Time spent palpating has been questioned as a measure of CBE performance due to differences in efficiency among practitioners of
different experience levels. However, it is valid to use this metric in this study because all participants in this study had the same experience level in CBE [4].

Affective performance was evaluated by having clinicians and medical education experts rate each of the three critical moments using a 5-item questionnaire. The questionnaire and data are presented in Appendices B and C. This questionnaire consisted of one item assessing the appropriateness of the learner’s response to the critical moment (would the response be appropriate with a real patient) and four items assessing the empathic quality of the learner’s response (does it encourage the patient to express emotion, validate the patient’s feelings, explore the patient’s feelings, and include appropriate nonverbal behavior). The four items dealing with empathy are taken from the empathy subscale of a validated instrument for evaluating medical interview skills [41].

4.2.3 Statistical Analysis

In presenting the results of this study, we refer to equivalence and non-inferiority. Equivalence indicates that the results of two treatments, while not identical, are so close that the treatments are equally preferable. However, equivalence is too restrictive for comparing MRIPS-CBE and SP as it does not allow MRIPS-CBE to outperform SP. For this reason, we additionally test for the non-inferiority of MRIPS-CBE. Non-inferiority of a novel treatment (MRIPS-CBE) indicates that the novel treatment is equally or more preferable than the existing treatment it is compared to (SP). Non-inferiority can be thought of as a lesser test of equality than equivalence. If two treatments are equivalent, each treatment is non-inferior compared to the other treatment. However, the reverse is not true, as non-inferiority does not imply equivalence.
Statistical equivalence is demonstrated by calculating a 95% confidence interval and comparing it to a clinically chosen zone of indifference. If the confidence interval lies completely inside the zone of indifference, the two treatments are equivalent. The zone of indifference is subjectively chosen based on knowledge of the scenario and measures used. In this experiment, many of the measures have a granularity of ±1 unit (e.g. one point on the 1-5 scales used in video coding; one item on the cognitive checklist). Thus we pick a zone of indifference to be a closed interval of length one: (-.5, .5). Non-inferiority is evaluated in a similar fashion but only the lower bound of the confidence interval must lie within the zone of indifference. The lower bound of this “one-sided” 95% confidence interval is equal to the lower bound of a standard 90% confidence interval, so in practice the 90% confidence interval (CI) is calculated.

To determine the validity of using MRIPS-CBE to evaluate learner performance, we look for a positive linear relationship between a participant's MRIPS-CBE performance and his SP performance. This is evaluated by calculating Pearson’s correlation coefficient.

All analysis was performed using the SPSS15 software package. In addition to evaluating equivalence and non-inferiority, we looked for effects of the order of treatments by performing a Student’s t-test between Group SP-MRH and Group MRH-SP.

4.2.4 Results and Discussion

4.2.4.1 Order effects

No effect of order was found on any measures. One might expect participants to exhibit improved performance in their second interaction. However, we did not expect
this to be the case. The two interactions were performed within an hour of each other and participants received no feedback as to the quality of their first performance. Thus, we did not expect learning and a corresponding improvement in the second interaction to occur.

4.2.4.2 Cognitive performance

Cognitive performance in MRIPS-CBE was significantly correlated with cognitive performance with the SP. This establishes the validity of MRIPS-CBE for evaluating learners’ cognitive performance in CBE.

One participant was removed from cognitive and affective measures analysis due to clearly demonstrating a disinterest in approaching the MRIPS-CBE interaction seriously. This participant first performed the SP interaction in which he queried 12 items on the cognitive checklist. However, in his MRIPS-CBE interaction he queried only 5 items on the cognitive checklist, only asking questions directly related to the manual exam in what appeared to be an attempt to complete the exam in a minimal amount of time. This was also the only participant to receive a rating of “1 – exceptionally inappropriate” from the medical expert video reviewers for his response to one of the critical moments. We chose to remove this participant from further analysis because we do not expect that this negative attitude towards MRIPS-CBE would be exhibited if MRIPS-CBE was integrated into a curriculum (i.e. if the participant was graded on his performance with MRIPS-CBE). Accordingly, we attach to the results the caveat that learners must approach the MRIPS interaction seriously in order to benefit. It is our opinion that this caveat extends to all educational experiences.

In the medical history checklist measure, the remaining nine participants queried an average of 10.3 ± 1.8 items from the MRIPS-CBE virtual human patient and queried
an average of 9.7 ± 2.2 items from the SP. A significant correlation between the number of items queried in MRIPS-CBE and SP was found, with $r(7) = 0.86 \ (r^2=0.74), p = 0.003$, significant at the $\alpha < 0.01$ level. This correlation shows that performance in a medical history of MRIPS-CBE is predictive of performance in a medical history of an SP. Johnsen et al. previously showed that medical student performance in a history-taking interpersonal simulation predicted performance in taking a medical history of an SP [81]. We expand on this approach further by evaluating non-inferiority of MRIPS-CBE.

Participants conducted more complete medical histories of the virtual human patient in MRIPS-CBE than of the SP, but not significantly so (by paired t-test: $t = 1.8, ns.$). The 90% CI is [-0.19, 1.53], demonstrating that MRIPS-CBE is non-inferior to an SP for practicing the cognitive task of medical history taking. This finding goes beyond the work of Johnsen et al., to demonstrate that for practicing taking a medical history a MRIPS-CBE interaction is an equally preferable substitute for an SP.

All participants arrived at the same differential diagnoses and diagnostic workups with MRIPS-CBE and the SP. In other words, if the participant arrived at the correct diagnosis for the SP, he also arrived at the correct diagnosis with MRIPS-CBE. Thus for diagnosis and diagnostic workup, participants’ performances in MRIPS-CBE and SP were equivalent and also perfectly correlated (i.e. $r^2 = 1.0$). Participants performed well in these two tasks, as 9 of 10 participants arrived at correct differential diagnoses and diagnostic workups.
These results demonstrate the validity of MRIPS-CBE for practice and evaluation of learners’ skills in the cognitive tasks of taking a medical history and integrating findings into a correct diagnosis.

4.2.4.3 Psychomotor

There were no sensors present in the SP breast, so to compare performance with MRIPS and the SP, we employed measures previously used in evaluating performance with SPs and Mammacare silicone breast models. Psychomotor performance was measured as the completeness of palpation: masses found and time spent palpating. These measures have previously been used to measure completeness of palpation, and Hall et al. indicated a positive linear relationship between these measures and completeness [65][66]. The time spent palpating the patients’ breasts was significantly correlated ($r^2(8) = 0.64, p < 0.01$). The participant’s ability to find masses in the SP’s breast was also significantly correlated with the participant’s ability to find the masses in the MRIPS-CBE breast ($r^2(8) = 0.63, p < 0.05$). Five participants found all masses in both the MRIPS-CBE and SP breasts; three participants did not find the masses in either breast; and only one participant found the mass in the SP breast but failed to find the masses in the MRIPS-CBE breast. These correlations validate MRIPS-CBE for evaluating learners’ skill in the psychomotor task of palpation.

On average, participants spent more time palpating the MRIPS-CBE breast ($83.3 \pm 35.4$ seconds) than they did palpating the SP breast ($65.9 \pm 25.7$ seconds). This was true for 9 or the 10 participants, even those who performed the SP exam before the MRIPS-CBE exam. The difference in average palpation duration was significant by a paired t-test: $t(9) = 2.6, p < 0.05$. The MRIPS-CBE mannequin included an armpit which could also be palpated as part of the exam, while the SP breast did not. However, time
spent palpating the armpit was not included in this comparison. The longer palpation duration for the MRIPS-CBE breast may be explained by the virtual human occasionally looking at the participant during palpation. The virtual human’s face was in the participant’s visual field of view during palpation, and alternated looking at the back wall of the room (gaze aversion) and at the participant’s head position. In contrast, the SP’s face was not in participants’ field of view during palpation, and the SP tended to watch the palpation of the silicone breast the entire time. This allowed the SP to have context to answer participant questions such as “is this the mass you found?” Whether lengthened by meeting eyes with the virtual human or not, it is clear that participants put at least as much effort into palpating the MRIPS-CBE breast as the SP breast. For this reason, we accept MRIPS-CBE as valid as a substitute for the SP for practice of the CBE psychomotor task of palpation.

Because the breasts palpated in the SP and MRIPS-CBE exams were the same simulated (silicone) breast, the SP and MRIPS-CBE exams were more similar than they would have been if the SP exam was performed on a human breast. However, silicone breasts have been validated for evaluating and learning self breast exams and CBEs [4][65][66]. This prior finding allows us to conclude that MRIPS-CBE is valid for evaluating learners’ psychomotor skills in CBE.

4.2.4.4 Affective

Results of the video review of the three critical moments are presented in Table 4-2. Video review evaluated the appropriateness and empathic content of participants’ responses to the critical moments in MRIPS-CBE and SP. Ratings demonstrated that participants’ performances in MRIPS-CBE were either non-inferior or equivalent to their performances with the SP. Participants exhibited similar affective behaviors with the
virtual human as with the SP. This establishes the validity of MRIPS-CBE for practicing affective components of clinical breast examination.

However, no correlation was found between affective performance in MRIPS-CBE and with the SP. Although MRIPS-CBE and SP are equally preferable for practicing affective components of CBE, MRIPS-CBE is not validated for evaluating affective performance in CBE. We believe the lack of correlation between affective performances in MRIPS-CBE and SP is due to the subjective nature of evaluating empathy and appropriateness of the participants’ responses to the critical moments. Reviewer difficulties in rating affective performance are demonstrated by the lack of inter-rater reliability. Inter-rater reliability was not established, as calculated inter-rater reliability coefficients (see Appendix C) were less than 0.70 for every item in the video review survey [117]. This emphasizes a need for objective measures of affective performance, and motivates our later approach of detecting key-phrases associated with empathy within user responses to critical moments (described in Section 7.3.1 and used in the user study of Chapter 9).

The medical educators and experts who performed the video review noted that the overall use of empathy was low. Indeed, only 5 of 10 participants were rated positively on expressing empathy in response to the virtual human’s “could this be cancer?” critical moment, and no participants were rated positively on expressing empathy in response to the virtual human’s “wait! I’m scared!” critical moment. Performance was actually worse for the SP critical moments (2/10 and 0/10 respectively), though this may be due to video reviewers’ increased expectations of performance with the SP. This low level of empathy likely contributes to the lack of correlation between MRIPS-CBE and
SP affective performance. However, this low level of expressed empathy is expected behavior of novice students. According to Pugh et al., novice learners of CBE perform poorly on affective and cognitive components until they master the psychomotor and cognitive-psychomotor components (e.g. palpation) of the exam [38]. This observation motivates the incorporation of real-time visual feedback for guidance of psychomotor, cognitive, and affective performance (Chapters 6-8) and subsequent evaluation of learning of all three skill sets in MRIPS-CBE (Chapter 9).

4.2.5 Limitations of the Study

A small population was obtained for this study, allowing us to find only large-size effects. Similarly to statistical tests of difference, results of statistical equivalence and non-inferiority tests will hold with larger population sizes. However, it is likely that a medical school considering curriculum integration of a simulation such as MRIPS-CBE would desire a similar study be run with a larger population.

The results of this study are only applicable to populations of novice learners of CBE, as all participants were nearing the end of their first year. This 1st-2nd year medical student group is the target end-user group for MRIPS-CBE. However, these novice learners performed poorly on cognitive and affective aspects of the exam with both MRIPS-CBE and the SP: on average asking only 10 of the 20 medical history questions and generally receiving negative (< 3.0) ratings on use of empathy in the critical moments. This poor performance is expected of novice learners. However, more experienced learners may not approach MRIPS-CBE with the similar positive attitude with which these novices approached both MRIPS-CBE and the SP. For this reason, equivalence and non-inferiority should be retested with more experienced learners if MRIPS-CBE were to be incorporated into, e.g., 3rd-year curriculum.
In this study, we wished to evaluate learners’ interactions with the virtual human and their ability to perform a CBE using the haptic interface of MRIPS-CBE. For these reasons, if speech recognition failed to produce reasonable text input from the participant’s utterance, the experimenter triggered the virtual human to reply appropriately to the participant. The result is a MRIPS interaction with idealized speech interface performance. It is possible that this artificially improved learners’ performance in the cognitive task of history taking. It is unlikely that it significantly impacted affective performance, as the critical moments, other than the moment concerning the patient’s mother, were triggered from manipulation of the haptic interface. The correlation between MRIPS and SP performance would not be impacted by a reduction in speech interface performance, provided all participants experienced a similar reduction in speech interface performance in MRIPS. However, this approach potentially impacted the result of equivalence or non-inferiority. We do not view this as impacting the conclusions of this study. The focus of this dissertation is not on speech interfaces, and we expect speech recognition and understanding to continue to improve, shrinking the gap between typical and ideal speech interface performance. Even using the current speech interface, performance can be improved through a training session with speech recognition. This training session was omitted from this study because of time constraints, but could be incorporated if MRIPS was used in a curriculum.

4.2.6 Conclusions

This study provided statistical evidence that for practicing the cognitive, psychomotor, and affective components of CBE, MRIPS-CBE is equally preferable as an SP. This establishes the use of MRIPS-CBE as an additional or even alternative tool for medical educators to give students more practice opportunities in CBE. However,
before MRIPS-CBE can be integrated into a curriculum in which students are graded, we must establish validity of MRIPS-CBE for evaluating a learner’s performance.

Study results validated MRIPS-CBE for evaluating cognitive and psychomotor performance in CBE, but did not establish validity for evaluating affective performance. The next study, Study MRIPSx2, employs another accepted technique [42] for establishing validity of a simulation for evaluating real-world skill. Study MRIPSx2 seeks to determine whether MRIPS-CBE can distinguish between participants of varied (known) skill and experience levels on dimensions of cognitive, psychomotor, and affective performance in CBE.

4.3 MRIPSx2

In Study MRIPSx2, participants of varied experience levels with CB examined a mixed reality human patient using MRIPS-CBE. If performance with MRIPS-CBE reflects users’ experience level with CBE, MRIPS-CBE will be validated for evaluating learners’ CBE skill. This approach has previously been used by Balkissoon et al. to establish the validity of a rectal exam simulator for evaluating learners’ rectal examination skills [42]. Although we have already established validity for evaluating cognitive and psychomotor skills, we will again evaluate users in all three skill sets: cognitive, psychomotor, and affective.

4.3.1 Study Design and Procedure

The study was conducted in two stages, in late May and early July 2008, in order to capture 2nd, 3rd, and 4th-year medical students as well as interns, residents, and clinicians. Forty-two medical students, residents, and clinicians at the Medical College of Georgia enrolled in the study. The population breakdown is presented in Table 4-3.
In the May 2008 study, participants completed a background survey to assess experience in CBE and then performed a CBE using MRIPS-CBE. The July 2008 study participants performed CBEs of two mixed reality human patients using MRIPS-CBE. In between the two MRIPS-CBE experiences, participants completed a review session in which they reviewed their CBE from the perspective of the patient. The analysis presented here focuses only on the first MRIPS CBE performed by participants.

4.3.2 Measures

Measures once again focus on cognitive, psychomotor, and affective components of the CBE. The measure of cognitive performance is the same medical history checklist used in Study MRIPS-SP, shown in Table 4-1. Data collected with this checklist is shown in Appendix F.

The psychomotor component evaluated was the completeness of palpation of the breast. Study MRIPSx2 marked the introduction of an improved version of the MRIPS-CBE mannequin, containing 64 force sensors (42 in the breast). This improvement (over the 12 force sensors in the breast used in Study MRIPS-SP) allowed us to use a more sensitive, precise measure of palpation completeness: the percent of breast tissue palpated. The location of each sensor was labeled by hand in an image of the mannequin. The area of tissue in which each sensor could detect force was estimated manually by observing sensor values while applying force to the breast in the pattern of a grid of approximately 1” x 1” squares. The area in which each sensor detected force was modeled as a circle. During the exam, if a sensor reported a value indicating palpation (light pressure or greater), the circle corresponding to that sensor was included in the area measured. The output generated by a participant’s exam is shown in Figure 4-3. The percent area palpated by each participant is given in Appendix G.
We have since developed an automated, more precise method for determining the area of tissue palpated as well as the force palpated with; this method is presented in Section 7.4.

To measure affective performance, we once again had medical experts review video of critical moments in participants’ exams. Three critical moments were included in the scenario used in this study. The first critical moment occurred when the participant asked about a family history of breast disease or cancer. The patient indicated that her mother had passed away recently from cancer: “I lost my mom to cancer two years ago.” The second critical moment was triggered by the participant’s action of opening the physical hospital gown of the MRIPS-CBE haptic interface. This occurred as the participant was about to begin palpating the patient’s breast. The patient exclaimed “Wait! I’m scared. What if you find cancer?” The third critical moment occurred after palpation when participants instructed the patient that she needed a mammogram. The patient asked “Do I really have to get a mammogram? I mean, my mom was fine, then she had a mammogram and all of the sudden she was really sick?”

The instrument used by the video reviewers was simplified from Study MRIPS-SP, in an effort to reduce the inter-rater variability observed in Study MRIPS-SP. The new instrument contained the same item querying appropriateness but reduced the four empathy items to a single item. The instrument is presented in Appendix A.2 and the ratings of participants in Appendix A.3. Video review was conducted only for the 19 participants who completed the study in July 2008.

4.3.3 Analyzing the Impact of Experience on Performance

Similar prior work in medical simulation has taken the approach of including participants of a wide variety of experience levels and performing post-hoc analysis to
find a grouping of experience levels which were discriminated between by the simulator (i.e. the analysis looked for two or more groups with significantly different performance using the simulator). For example, in evaluating validity of a rectal exam simulator, Balkissoon et al. started with participants in three groups: medical students, residents, and clinicians. However, the final analysis included only two groups based on the number of rectal exams previously performed: low experience (< 5) and high experience (>= 5 exams) [42].

Because we analyze three distinct aspects of performance, we considered different groupings of experienced and inexperienced participants for each of the cognitive, psychomotor, and affective performance.

Affective performance consisted of comforting a fearful patient and expressing empathy concerning the patient’s loss of her mother to breast cancer. Expressing empathy and comforting a patient are skills which are learned in a variety of scenarios. Thus, experience with CBE is not necessarily a good predictor of affective performance. Instead, overall experience with human patients is expected to be a better criterion for classifying affective performance. Consulting with medical educators at the Medical College of Georgia, we grouped participants into two groups to compare affective performance. The affective-experienced group contained residents and clinicians and the affective-inexperienced group contained medical students and interns.

Psychomotor performance in CBE can only be improved through performing CBEs [4][65]. In evaluating psychomotor performance, we grouped participants based on their experience in CBE of human patients. This yielded two groups. The psychomotor-experienced group contained participants who had completed a womens’ health
clerkship including having their CBE evaluated by an expert in clinic. The psychomotor-
inexperienced group had not yet begun this clerkship. This classification corresponded
with the <5 exams vs. >= 5 exams classification used by Balkissoon [42]. In this
analysis, we included only medical students and interns, as these participants had
received training in CBE at the Medical College of Georgia. This institution teaches the
Mammacare method of examination (circular palpations, vertical strip pattern, three
levels of pressure) [65]. The Mammacare method is known to be the most effective
method for finding masses, but many other methods exist [4]. By restricting analysis to
participants who have been taught the Mammacare method we avoided incurring
additional variance due to the method of examination used.

Cognitive performance was measured using the medical history checklist of Table
4-1. This checklist focuses on medical history items related to assessing breast health
history and breast cancer risk factors. For this reason, we expect experience in CBE to
be a predictor of performance in this cognitive task. Thus the cognitive-experienced
and cognitive-inexperienced groups used the same classification criteria as the
psychomotor groups. As this measure assesses completeness of the history as the
number of items addressed in the medical history, we omitted clinicians and residents
from this analysis. It is known that tactics in taking a medical history change as
advanced levels of experience are reached [12]. The more experienced residents and
clinicians are likely to use a smaller, more focused set of questions and infer information
from their past experiences or omit a question if the same information can be obtained
in the physical exam (e.g. omitting asking about nipple discharge because nipple
discharge can be tested in the exam). Indeed, use of checklists in SP encounters has
been found to be biased against these more efficient clinicians or experts [118]. Expert behavior is in contrast to behavior we have observed with medical students, who tend to stick to a “script” of questions they have learned in lecture. Thus the number of relevant questions asked appears to be a good indicator of student skill and performance, but not a good indicator of highly experienced residents and clinicians.

4.3.4 Results

Results are summarized in Tables 4-4 through 4-6. Performance in MRIPS-CBE distinguished between experienced and inexperienced students in cognitive, psychomotor, and affective tasks. This further establishes the validity of MRIPS-CBE for evaluating learners in these three skill sets.

4.3.4.1 Cognitive

The experienced medical students and interns who had received expert tutelage in the women’s health clerkship asked an average of three more questions (from the medical history completeness instrument, Table 4-1) than the inexperienced medical students (Table 4-4). Experienced students asked 12.3 ± 2.3 critical questions compared to 9.3 ± 2.9 for the inexperienced students. This difference was significant by an independent samples t-test: t(26) = 2.9, p < 0.01. This result provides evidence in addition to that of Study MRIPS-SP to validate MRIPS-CBE for evaluating the cognitive components of learners’ CBES.

4.3.4.2 Psychomotor

Experienced medical students and interns with five or more prior CBES performed significantly more complete examinations of the breast than did inexperienced medical students with less than five prior CBES. Experienced participants palpated an average of 84.2% ± 7.6% of the breast tissue with light or higher pressure, while inexperienced
participants palpated 74.8% ± 13.4% of the breast tissue. The difference was significant at $\alpha < 0.05$ by independent samples t-test: $t(30) = 2.4$, $p = 0.03$. This result provides evidence in addition to that of Study MRIPS-SP to establish the validity of MRIPS-CBE for evaluating the psychomotor component of learners’ CBES.

4.3.4.3 Affective

Six medical educators and clinicians evaluated participants’ affective performances in three critical moments (see Section 4.3.2) using the instrument of Appendix A.2. As in Study MRIPS-SP, inter-rater reliability was not above “adequate” agreement indicated by a coefficient value of 0.70 [117]. Considering that adequate inter-rater reliability was not achieved using the previously validated instrument used in Study MRIPS-SP nor with the simplified instrument used in this study, it appears as though rating of empathy is too subjective to achieve high inter-rater reliability.

The six expert reviewers’ scores were averaged to produce a score of acceptability and empathy for each participant’s response to each of the three critical moments (Table 4-6). The experienced group outperformed the inexperienced group in both acceptability of response and empathic content of response for all three critical moments. However, the difference in appropriateness was significant for two out of three moments and the difference in empathy was significant for only the critical moment concerning the patient’s fear of having a mammogram. Given better inter-rater agreement, we expect that the amount by which the experienced participants outperformed the inexperienced participants will become significant. Without inter-rater agreement we still accept that MRIPS-CBE is able to distinguish between experienced and inexperienced affective performance, as the mammogram critical moment produced
significantly different performance based on experience. This result establishes the
validity of MRIPS-CBE for evaluating the affective component of learners’ CBEs.

4.3.5 Discussion

The results of this study establish the validity of MRIPS-CBE for evaluating
learners’ cognitive, psychomotor, and affective performance. As in Study MRIPS-SP,
the experimenter triggered virtual human responses when speech recognition failed.
This was performed uniformly across participants, so it does not affect the statistical
results of the study.

4.4 Conclusion

Study MRIPS-SP established that users of MRIPS-CBE use their real-world skills
to perform CBEs that are similar to CBEs of SPs along cognitive, psychomotor, and
affective dimensions. Study MRIPSx2 established that experienced participants are
able to perform more complete and correct CBEs than inexperienced participants, along
cognitive, psychomotor, and affective dimensions. These two studies established the
validity of MRIPS-CBE as a tool for practicing CBEs and as an instrument for evaluating
learners’ cognitive, psychomotor, and affective skills in CBE.

From a computer science or human-computer interaction perspective, the main
result of these studies is to provide evidence that users of mixed interpersonal
simulation treat their mixed reality human interaction partners similarly to how they treat
human interaction partners.

From a simulation perspective, the main result of these studies is to show that an
interpersonal simulator incorporating haptic interfaces elicits and can be used to
evaluate users’ real-world cognitive, psychomotor, and affective skills in an
interpersonal scenario.
From a medical education perspective, the main result of these studies is to establish MRIPS-CBE as an additional practice opportunity for learners of CBE. Since current medical curricula do not provide enough practice opportunities for many learners to reach competence in CBE [9], MRIPS-CBE is a viable addition to a medical curriculum for teaching CBE. MRIPS-CBE can be used to augment an existing SP curriculum and has been shown to be a valid substitute for an SP interaction.

Prior medical simulators have been incorporated into medical curricula as a result of establishing validity for evaluating learner skill (e.g. Pugh’s breast palpation, pelvic, and rectal exam trainers into Northwestern University’s curriculum [5][38][42]). However, prior interpersonal simulations, for which content validity has been established, have not been accepted into interpersonal skills curricula (e.g. the medical history simulator of Johnsen et al [10] is not currently used in the curricula of either medical school which assisted in its development). One of the driving goals of this work is to establish that mixed interpersonal simulation is an invaluable addition to a curriculum in which the simulated interpersonal scenario is currently underserved (e.g. CBE education in current medical curricula [9]).

For this reason, we continue beyond establishing the validity of MRIPS-CBE, to investigate learning and training transfer in MRIPS-CBE (Chapter 9). Before evaluating learning and training transfer, we incorporate real-time and post-experiential feedback of learner performance (Chapters 6-8). Such feedback is mandated for learning in interpersonal scenarios [12].
Table 4-1. Items in the medical history completeness checklist used to evaluate participants’ cognitive performance.

<table>
<thead>
<tr>
<th>History of Present Illness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of current complaint (for MRIPS, pain; for SP, breast mass)</td>
</tr>
<tr>
<td>Location of current complaint (for MRIPS, pain; for SP, breast mass)</td>
</tr>
<tr>
<td>Presence of discharge</td>
</tr>
<tr>
<td>Other breast changes (e.g. redness, skin puckering)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medical History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menarche: age of onset</td>
</tr>
<tr>
<td>Menarche: currently occurring or post-menopausal</td>
</tr>
<tr>
<td>Use of hormones for birth control or hormone replacement</td>
</tr>
<tr>
<td>Past pregnancies</td>
</tr>
<tr>
<td>Past breast problems</td>
</tr>
<tr>
<td>Screening: yearly clinical breast exams by a doctor</td>
</tr>
<tr>
<td>Screening: monthly self breast exams</td>
</tr>
<tr>
<td>Screening: yearly mammograms</td>
</tr>
<tr>
<td>Past hospitalizations</td>
</tr>
<tr>
<td>Past surgeries</td>
</tr>
<tr>
<td>Current medications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family history of cancer</td>
</tr>
<tr>
<td>Other family history of medical problems</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking / use of tobacco</td>
</tr>
<tr>
<td>Drinking / use of alcohol</td>
</tr>
<tr>
<td>Health risks involved in employment</td>
</tr>
</tbody>
</table>

Table 4-2. Results of video review of critical moments Study MRIPS-SP (scores are averages of the multiple experts’ ratings).

<table>
<thead>
<tr>
<th></th>
<th>MRIPS</th>
<th>SP</th>
<th>CI</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Wait! I’m scared. What if you find something bad?”</td>
<td>Appropriate</td>
<td>3.92 ± 0.39</td>
<td>3.48 ± 0.30</td>
<td>[0.10, 0.78]</td>
</tr>
<tr>
<td></td>
<td>Empathic</td>
<td>1.30 ± 0.17</td>
<td>1.28 ± 0.007</td>
<td>[-0.10, 0.15]</td>
</tr>
<tr>
<td>“Can you tell me what happens next?”</td>
<td>Appropriate</td>
<td>3.89 ± 0.49</td>
<td>3.24 ± 0.85</td>
<td>[0.28, 1.02]</td>
</tr>
<tr>
<td></td>
<td>Empathic</td>
<td>3.10 ± 0.93</td>
<td>2.95 ± 0.81</td>
<td>[-0.40, 0.71]</td>
</tr>
</tbody>
</table>
Table 4-3. Population breakdown for Study MRIPSx2.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Population</th>
<th>Study stage</th>
<th>Educational experience in CBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical student Yr. 2</td>
<td>5</td>
<td>May</td>
<td>Lecture</td>
</tr>
<tr>
<td>Medical student Yr. 3</td>
<td>12</td>
<td>May, July</td>
<td>Women's health clerkship with real patients</td>
</tr>
<tr>
<td>Medical student Yr. 4</td>
<td>12</td>
<td>May, July</td>
<td>Clinic (real-patients)</td>
</tr>
<tr>
<td>Intern year</td>
<td>3</td>
<td>July</td>
<td>Clinic (real-patients)</td>
</tr>
<tr>
<td>Resident</td>
<td>5</td>
<td>July</td>
<td>Clinic (real-patients)</td>
</tr>
<tr>
<td>Clinician</td>
<td>5</td>
<td>July</td>
<td>Clinic (real-patients)</td>
</tr>
</tbody>
</table>

Table 4-4. Cognitive performance results for Study MRIPSx2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pop. Size</th>
<th>Mean ± stdev</th>
<th>95% CI</th>
<th>Test</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced*</td>
<td>12</td>
<td>12.3 ± 2.3 items</td>
<td>[10.8, 13.7]</td>
<td>t-test, t(26) = 2.9</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Inexperienced**</td>
<td>16</td>
<td>9.3 ± 2.9 items</td>
<td>[  7.8, 10.9]</td>
<td>t(26) = 2.9</td>
<td></td>
</tr>
</tbody>
</table>

(*) Post-women’s health clerkship; (**) Pre-women’s health clerkship.

Table 4-5. Psychomotor performance results for Study MRIPSx2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pop. Size</th>
<th>Mean ± stdev</th>
<th>95% CI</th>
<th>Test</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced*</td>
<td>14</td>
<td>84.2% ± 7.5%</td>
<td>[79.9, 88.6]</td>
<td>t-test, t(30) = 2.4</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Inexperienced**</td>
<td>18</td>
<td>74.8% ± 13.4%</td>
<td>[68.1, 81.4]</td>
<td>t(30) = 2.4</td>
<td></td>
</tr>
</tbody>
</table>

(*) Post-women’s health clerkship; (**) Pre-women’s health clerkship.
Table 4-6. Affective performance results for Study MRIPSx2 (scores are averages of multiple experts’ ratings).

<table>
<thead>
<tr>
<th>Group</th>
<th>Pop. Size</th>
<th>Mean ± stdev.</th>
<th>Test statistic</th>
<th>Sig.</th>
</tr>
</thead>
</table>
| 1. “Wait! I’m scared. What if you find cancer?”
  a. Appropriate?        | 8         | 3.9 ± 0.7     | t(25) = 0.5    | ns   |
  Experienced*           | 8         | 3.8 ± 0.4     |                |      |
  Inexperienced**        | 19        | 2.5 ± 1.1     | t(25) = 0.8    | ns   |
  b. Empathetic?         | 8         | 2.5 ± 1.1     |                |      |
  Experienced*           | 8         | 2.4 ± 0.9     |                |      |
  Inexperienced*         | 19        | 2.4 ± 0.9     |                |      |

2. “I lost my mother to breast cancer two years ago.”
  a. Appropriate?
  Experienced*           | 9         | 3.9 ± 0.5     | t(23) = 2.2    | p < 0.05 |
  Inexperienced**        | 16        | 3.4 ± 0.6     |                |      |
  b. Empathetic?
  Experienced*           | 9         | 2.3 ± 0.9     | t(23) = 0.9    | ns   |
  Inexperienced*         | 16        | 2.0 ± 0.9     |                |      |

3. “Do I really have to get a mammogram…”
  a. Appropriate?
  Experienced*           | 8         | 4.2 ± 0.3     | t(21) = 2.4    | p < 0.05 |
  Inexperienced**        | 15        | 3.7 ± 0.5     |                |      |
  b. Empathetic?
  Experienced*           | 8         | 3.3 ± 0.9     | t(21) = 2.6    | p < 0.05 |
  Inexperienced*         | 15        | 2.2 ± 0.9     |                |      |

(*) Residents and clinicians; (**) 2\textsuperscript{nd}-4\textsuperscript{th} year students and interns.

Figure 4-1. Appearance of the MRIPS-CBE patient for Study MRIPS-SP. Participants wore a head-mounted display to view the virtual human.
Figure 4-2. The appearance of the MRIPS-CBE mixed reality human in Study MRIPSx2. The user’s view in the HMD is shown as the outlined inset.

Figure 4-3. Visualization of a participant’s CBE completeness. Green indicates tissue palpated and red indicates tissue missed in the exam.
This chapter describes the development of MRIPS-NEURO, an interpersonal simulation of a neurological examination with abnormal findings of cranial nerve palsy. An early version of the system was published in the proceedings of the IEEE Symposium on 3D User Interfaces 2009 [28]. The design of the system and ability to simulate the neurological exam has since been significantly enhanced.

**Collaborators:** Kyle Johnsen developed the original eye movement model, virtual human abilities such as counting the user’s fingers and reading from an eye chart, and a portion of the tool interaction. The appearance of the virtual human and tools used in the exam was developed by Brent Rossen, Kyle Johnsen, and me. Medical collaborators Juan Cendan, Bayard Miller, Lou Ritz, and Thea Nalls provided information on neurological exam content and feedback concerning the correctness of the abnormal eye movements.

**Personal contributions:** I developed a new eye movement model, integrated touch into the interaction, developed virtual human nonverbal actions to support additional neurological exam tests, and developed tool use such as the fundoscopic exam.

**Relevance to thesis:** The thesis states that interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance improves users’ cognitive, psychomotor, and affective performance in an interpersonal scenario. To prove this statement, we must develop such an interpersonal simulation and evaluate the impact of feedback on learner performance within this interpersonal simulation. This chapter describes the design of a mixed reality
interpersonal simulation incorporating haptic interfaces and providing real-time feedback (Chapter 8). The simulation described in this chapter, MRIPS-NEURO, is used to evaluate the impact of this real-time feedback on learners' cognitive, psychomotor, and affective performance (Chapter 8).

5.1 Introduction

A mixed reality interpersonal simulation, MRIPS-NEURO, was developed to simulate a neurological exam with abnormal physical findings, e.g. abnormal eye movements. The motivation for developing this simulation was to provide a learning opportunity not afforded by current medical education and simulation approaches. MRIPS-NEURO affords communication and physical examination of a life-sized virtual human agent through speech and touch. MRIPS-NEURO also introduces a new affordance to MRIPS: hand-held tool use and gestures through the manipulation of the haptic interface. We have developed ophthalmoscope, eye chart, and gesture (appearing as the user’s right hand) tools for communication and exam performance.

Using MRIPS-NEURO, Figure 5.1, novice learners of neurological exams can practice history taking (verbally evaluating non-visible symptoms such as headache), physical exam tasks (eliciting abnormal physical findings), synthesizing the information collected in history taking and physical exam into a differential diagnosis of the patient's abnormality, and practice affective exam components such as addressing the patient's concerns and issues of patient safety.

In addition to the science-oriented advances of incorporating hand-held tool use and hand gestures into the interaction with the virtual human, MRIPS-NEURO incorporates an engineering-oriented advance: creation of a virtual human that is able to simulate cranial nerve abnormalities.
As with MRIPS-CBE, the design of MRIPS-NEURO is divided into three parts: the appearance of the virtual human and virtual world, the underlying virtual human simulation, and the haptic interface. The virtual human simulation of MRIPS-NEURO operates similarly to the simulation module of MRIPS-CBE described in Chapter 3, so we omit a full treatment from this chapter. This chapter provides background information concerning the neurological exam, then focuses on the development of a virtual human capable of presenting the physical findings necessary for the exam, and on the incorporation of a haptic interface for performing the physical exam.

5.1.1 The Neurological Exam Requires Cognitive, Psychomotor, and Affective Skills

As with the CBE, the neurological exam is composed of cognitive, psychomotor, and affective components.

The cognitive parts of the cranial nerve exam are:

- Recall of important questions to ask in order to compile a medical history and assess symptoms that are not physically expressed, e.g., onset of vision problems, headache [47].

- Recall of physical exam tasks that should be used to evaluate disorders of the patient’s 12 cranial nerves.

- Synthesis of information gathered from the medical history and the cranial nerve tests into a differential diagnosis – determining what nerve is affected.

The synthesis of information is known to be the most difficult component of the exam, as there are many tests and a dozen cranial nerves, and learners do not have experience conducting a physical exam with abnormal findings alongside taking a medical history of a patient (because learners do not have opportunities to examine peers, SPs, or real patients with abnormal findings).
Many of the tests used in the physical exam involve manipulation of hand-held tools or gestures, and could be thought of as psychomotor tasks. However, educators do not identify the majority of these tasks as providing difficulties for learners. Thus only two tasks will be considered to be psychomotor tasks in this work.

One of the two psychomotor tasks is using the ophthalmoscope to conduct a fundoscopic exam. The ophthalmoscope is a complex instrument which requires significant practice to master. MRIPS-CBE does not aim to teach mastery of the ophthalmoscope, as this is better served by existing purely physical approaches such as practice with a peer or mannequin [56]. Instead MRIPS-CBE incorporates a simplified fundoscopic exam which focuses on the cognitive task of interpreting the appearance of the patient’s retina.

The second psychomotor task is testing the patient’s eye movements by sweeping one’s finger (or light) in the shape of an uppercase “H” while the patient follows the finger with his eyes. This task requires that the shape of the “H” and depth of the finger from the patient’s head be such that the extremes of the patient’s vision are examined (e.g. the patient must move his eyes as far left as possible). If the novice learner does not use an “H” that elicits the extremes of eye movement, the learner misses information that may indicate an abnormality with one of the cranial nerves, e.g. the patient’s inability to move an eye to one of the extremes.

In addition to these cognitive and psychomotor tasks, the learner can also practice affective components of the exam. As with the CBE, the patient is often anxious and it is the learner’s job to keep the patient comforted through appropriate expressions of empathy. Unique to the neurological exam are concerns of the patient’s immediate
safety. For example, did a patient with severe double vision drive to the clinic? Is it safe for him to drive home? A learner who is more concerned with a correct diagnosis than understanding what the patient is going through may disregard issues of safety. The patient’s safety is additionally the doctor’s responsibility and even liability [119], so it bodes well for a learner’s future to be exposed to issues of patient safety as a novice.

5.1.2 Evaluating MRIPS-NEURO

The evaluation portion of my dissertation focuses on two components: 1) evaluating the impact of real-time feedback on learner performance and 2) evaluating whether learning occurs in MRIPS and transfers to the real-world interpersonal scenario.

The first component is addressed in MRIPS-NEURO: the impact of real-time feedback of learners’ cognitive, psychomotor, and affective skills is evaluated in a formal study described in Chapter 8. However, the second component can not be fully evaluated in MRIPS-NEURO because a standardized real-world scenario is unavailable for evaluation of learning and training transfer. Neither SPs nor real patients with abnormal findings are available to evaluate learners’ skills in a real-world neurological exam. In fact, MRIPS-NEURO represents the first step towards a standardized platform for evaluating learners’ skills in the interpersonal scenario of a neurological exam with abnormal findings.

Without the capability to directly evaluate learning in MRIPS-NEURO, we focus on a necessary step towards learning: content validity. Establishing content validity answers the question: Does MRIPS-NEURO simulate abnormal physical findings and the performance of neurological exam tasks to a high-enough fidelity that novice
learners are able to practice diagnosing the cranial nerve disorder and arrive at a correct diagnosis?

Such a practice opportunity is not provided by current medical education and simulation approaches due to the inability to simulate abnormal findings. Thus demonstrating the content validity of MRIPS-NEURO would be a significant step towards providing medical educators and students with a platform affording practice and evaluation of the neurological exam scenario.

To establish the content validity of MRIPS-NEURO, we conducted a user study which focused on whether medical students (2nd and 3rd year) were able to arrive at a correct differential diagnosis through performing an examination in MRIPS-NEURO. Novice learners have previously learned the mapping of symptoms and abnormal physical appearances to disorders of specific cranial nerves (from lectures and books). To establish content validity, we needed to show that these novice learners could use MRIPS-NEURO to perform all of the tests necessary for the learner to collect enough information to arrive at a correct diagnosis. These tests included manipulating hand-held tools, gestures, and communicating with the virtual human. Twelve of fourteen participants were able to arrive at a correct diagnosis; this was a proportion significant beyond chance, by a one-way Chi-square test ($X^2 = 5.8, p = 0.02$), establishing the content validity of MRIPS-NEURO. This study is described further in Section 5.5.

5.2 The Neurological Exam

In this work we use the term neurological exam to refer to a focused exam of the cranial nerves. There are twelve cranial nerves; MRIPS-NEURO focuses on simulating tests used to examine cranial nerves 2, 3, 4, 5, 6, 7, and 12. The functions of these nerves and symptoms resulting from disorders of these nerves are listed in Table 5-1.
Examining this subset of nerves requires manipulation of an ophthalmoscope with a light, an eye chart, and verbal and nonverbal communication involving hand gestures and references to tools. The virtual human developed for MRIPS-CBE complains of experiencing double vision. The primary tasks of an exam to evaluate double vision are listed in Table 5-2. The virtual human’s abilities to support these tasks and other components of a neurological exam are discussed in Section 5.3.2.

5.3 A Virtual Human Agent to Simulate Cranial Nerve Disorders

We developed a virtual human agent capable of displaying the physical symptoms of cranial nerve 3 and cranial nerve 6 palsies and capable of performing many of the tasks asked of the patient in a neurological exam of cranial nerves 2, 3, 4, 5, 6, 7, and 12. The design of this virtual human is broken into two sections: developing a model of eye movements that presents physiologically accurate abnormalities and designing the virtual human to perform tasks used to diagnose cranial nerve disorders.

5.3.1 Eye Movement Model

The most important aspect of simulating a cranial nerve disorder is displaying correct abnormal eye movements. These eye movements are the primary basis for diagnosis, in addition to other aspects of the patient’s appearance such as tilting of the head indicating CN4 palsy and secondary aspects such as headache [120] or trauma to the head.

In MRIPS-CBE the virtual human’s left eye is the eye affected by the CN3 or CN6 disorder; the right eye retains normal movement. Both normal and abnormal movements are controlled by the same model.

Moving the human eye to look at an object or perform a task such as following the doctor’s fingers requires a complex interaction of six muscles. The original eye
movement model, developed by Kyle Johnsen, sought to replicate the effects of each muscle on the yaw and pitch of the eye [28]. However, this approach only modeled the primary function of each muscle, making it unable to simulate physiology such as greater angle of pitch when looking directly up than when looking up and to the right. Additionally, this approach could not simulate abnormalities related to cranial nerve palsies such as having the affected eye look down and out when the virtual human is attempting to look straight ahead.

For these reasons, a new eye movement model was developed that is not physically based, but does produce output consistent with real physiology. This model uses linear interpolation of the eight cardinal eye positions to restrict eye movement in a way that appears physiologically correct for the cranial nerve disorder. The eight cardinal eye positions are displayed in Figures 5-2 through 5-4 for no CN disorder, CN3 palsy, and CN6 palsy respectively. Typically “up” and “down” are not considered cardinal positions, but they are included here. Through review of case data and textbook diagrams [60], the UC Davis eye simulator [57], and discussion with medical collaborators, we estimated the yaw and pitch of each of the cardinal poses (and the default, looking straight pose) for normal eye movement, CN3, CN4, and CN6 – these are the three cranial nerve palsies which result in abnormal eye movement.

For each eye, the movement model defines a set of eight vectors corresponding to the cardinal movements, each 45 degrees apart. Each vector is associated with a (yaw, pitch) pair representing the maximum yaw and pitch of the eye for that cardinal position. If we think of the Cartesian plane as having dimensions of yaw and pitch, these vectors divide the plane into eight sectors. This is illustrated in Figure 5-5.
Given a desired gaze position (e.g. “look to the left”, “follow my finger”, or gazing at the user’s head position), this model outputs a new gaze position altered by the constraints imposed on eye movement by the cranial nerve disorder. For an affected eye, the process proceeds as (illustrated in Figure 5-5):

1. Given the desired gaze position, calculate the (yaw, pitch) required to rotate the eye from its default (straight ahead) position.
2. This (yaw, pitch) pair defines a vector \( \mathbf{d} \), the desired gaze vector.
3. Assuming \( \mathbf{d} \) originates from the origin of the normal eye model, determine which two normal eye movement vectors \( \mathbf{v_1}, \mathbf{v_2} \) define the sector in which \( \mathbf{d} \) lies.
4. Calculate the angles between \( \mathbf{d} \) and \( \mathbf{v_1}, \mathbf{v_2} \) as \( a_1 \) and \( a_2 \).
5. Normalize and invert \( a_1 \) and \( a_2 \):
   a. \( a_1' = a_1 / (a_1 + a_2) \);
   b. \( a_2 = 1.0 - a_1' \);
5. \( a_1 \) and \( a_2 \) are now weights for linear interpolation. The smaller the angle between \( \mathbf{d} \) and the neighboring vector (\( \mathbf{v_1} \) or \( \mathbf{v_2} \)), the larger the weight.
6. Using the abnormal vectors \( \mathbf{v_1}' \) and \( \mathbf{v_2}' \) corresponding to the \( \mathbf{v_1} \) and \( \mathbf{v_2} \), find the vector \( \mathbf{d}' \) defining the maximum allowed (yaw, pitch) along the desired gaze vector \( \mathbf{d} \): \( \mathbf{d}' = (\mathbf{v_1}'a_1 + \mathbf{v_2}'a_2) / (a_1 + a_2) \).
8. Desired gaze length \( l_d = ||\mathbf{d}|| \); Maximum gaze length \( l_m = ||\mathbf{d}'|| \)
9. If \( l_d > l_m \), set \( l_d = l_m \).
10. Let \( \mathbf{s} \) indicate the (yaw, pitch) vector corresponding to the abnormal eye looking straight ahead.
11. The final gaze vector \( \mathbf{g} \) is then calculated as: \( \mathbf{g} = \mathbf{s} + \mathbf{d}'( l_d / l_m ) \). Example vectors \( \mathbf{d} \) and \( \mathbf{g} \) are illustrated for the case of CN6 in Figure 5-5.

The same process is performed for the normal eye, with the substitution of normal eye vectors for the abnormal eye vectors used in Step 7 (i.e. \( \mathbf{v_1}' = \mathbf{v_1}; \mathbf{v_2}' = \mathbf{v_2} \)).

Once the final gaze yaw and pitch angles are calculated, the eye does not rotate to these angles instantly, but moves over time. To accomplish this we define the maximal angle the eye can rotate per second and linearly interpolate between the current eye yaw and pitch angles and the desired gaze yaw and pitch angles based on the maximal angle per second and the duration of the current frame.
We consulted with neurologists and eye movement experts at the University of Florida’s College of Medicine to establish the correctness of the eye movements produced by this model.

5.3.2 Virtual Human Abilities to Support Neurological Tests

In addition to displaying physiologically accurate eye movements, the virtual human has verbal and nonverbal behaviors to support additional neurological tests. The virtual human supports additional tests of cranial nerves 2, 3, 4, 5, 6, 7, and 12, with the following abilities:

- Pupillary response: The virtual human’s pupils constrict when the ophthalmoscope light is shined into an eye. In the CN3 affected eye, the pupil remains dilated even in the presence of light. This test requires manipulation of the ophthalmoscope tool using the haptic interface. (Tests CN 2, 3).

- Eye movement: The patient can hold his head still and follow the ophthalmoscope light or the doctor’s finger with his eyes. This test requires user speech and manipulation of the gesture or ophthalmoscope tool using the haptic interface. (Tests CN3, 4, 6).

- For CN3, the virtual human can be asked to hold his drooping left eyelid up to enhance the doctor’s view of the eye movement.

- The virtual human can verbally express whether he sees double depending on where his eyes are looking. For example, with CN6 the patient’s double vision goes away if he looks to the right.

- Visual acuity: The virtual human can read from an eye chart, either reading the smallest line he can make out, or reading a line that the doctor points to using the haptic interface. This test requires user speech and manipulation of the eye chart tool using the haptic interface (Tests CN2, 3, 4, 6).

- With both eyes open, the virtual human will not be able to read any lines with CN3, 4, or 6. The virtual human is able to cover either eye with his hand, allowing him to read the eye chart with 20/20 visual acuity.

- Facial sensation: The virtual human’s facial sensation can be tested by touching the virtual human’s face using the haptic interface’s gesture tool. If the eye is accidentally touched, e.g. by the ophthalmoscope during the fundoscopic exam, the virtual human blinks and jerks his head away as a reflex. (Tests CN 5).
Facial maneuvers: The virtual human can smile, frown, raise his eyebrows, puff out his cheeks, blink, and wink. The patient can also stick out his tongue. These are all elicited through verbal instructions from the user. (Tests CN 7 and 12).

5.4 The Haptic Interface

The haptic interface of MRIPS-NEURO takes a different approach than the haptic interface of MRIPS-CBE. Unlike MRIPS-CBE, the MRIPS-NEURO haptic interface does not provide a physical representation of the patient. Although touching the patient's face is part of the neurological exam, the exam predominantly consists of tool manipulation, verbal references to manipulated tools, and hand gestures. Thus the haptic interface for MRIPS-NEURO should focus on providing natural manipulation of hand-held tools and robust detection of hand gestures.

5.4.1 Prior Approaches

Prior attempts to incorporate natural manipulation of hand-held tools and hand gestures into mixed and virtual environments have taken two approaches: encumbering the user with gloves and wires or vision-based gesture recognition. Gloves containing bend and pinch sensors can provide low-noise recognition of gestures [121] suitable for this application. However gloves alone do not provide the feel and weight of hand-held tools. Most importantly, in our experience with over 500 end-users in the medical profession (at the University of Florida and Medical College of Georgia), requiring the user to wear encumberances such as gloves significantly decreases acceptability of the simulation and would be a significant hurdle to the incorporation of the simulation into a medical curriculum.

Unencumbered alternatives to gloves have focused on vision-based tracking of the user's hands and has afforded simple hand and gesture based interaction in VEs [122][123]. However the amount and complexity of recognizable gestures is suitable
only for simple interfaces such as those used for virtual environment navigation and selection (e.g. point, pinch), and the accuracy and update-rate of such systems would also reduce acceptability of the simulation.

Former collaborator Xiyong Wang developed a minimally encumbering (no wires or gloves) interface based on optical tracking of user-worn infrared fiducials. However the optical tracking proved to be too noisy for the system to be usable to conduct the neurological exam. Problems that are trivial for humans, such as distinguishing one finger from two fingers prove to be extremely difficult to accomplish using vision and optical techniques.

Because of the drawbacks of prior approaches, we chose to base our haptic interface around a hand-held tracked device, sacrificing the naturalness of gestures for noise-free gesture recognition, and preserving natural manipulation of hand-held tools.

5.4.2 Haptic Interface: Wii-Remote and External Sensing

The haptic interface for MRIPS-NEURO is a hand-held tracked input device with the weight and shape of many hand held tools, such as the ophthalmoscope used in the neurological exam, and button inputs to simulate affordances of real tools.

We chose a single hand-held device instead of having one device for each tool. This is motivated by the ability to overload the single hand-held device to simulate multiple tools and perform multiple tasks. In contrast, the approach of having many tool-specific devices requires switching between devices any time a new tool is desired, and requires significant development if a new tool is required [124]. The single hand-held device takes a Swiss-army-knife approach, affording the simulation of multiple hand-held tools using a single interface which is representative of a wide class of hand-held tools. Thus the hand-held tools in MRIPS-CBE are virtual and manipulated by a
hand-held device that serves as the physical haptic interface (providing both passive haptic feedback through shape and weight, as well as active-haptic vibratory force feedback).

The virtual tools of MRIPS-CBE are an ophthalmoscope, an eye chart, and the gesture tool which appears as the user’s right hand and fingers. The tools and their use in the exam are described in Section 5.4.3.

The hand-held device used in the haptic interface was chosen to be the Nintendo Wii-Remote (“wiimote”) which we augmented with external six degree-of-freedom tracking using a four camera Naturalpoint OptiTrack infrared tracking system.

The wiimote was chosen because it is shaped as a hand-held tool and has high degree of freedom control. The wiimote features an array of integrated sensors that are reported at 100Hz update rate over a Bluetooth connection: 11 buttons, 3 orthogonal accelerometers (+/- 3g), and a 45-degree field-of-view infrared camera (128x96) that tracks up to 4 points at 1/8 to 1 pixel resolution depending on the size of the infrared point. In addition, the wiimote can display information through integrated LEDs, speaker, and vibration motor. Open source software was used to acquire data from the wiimote across the Bluetooth interface [125]. A summary of the wiimote’s potential as an interaction device and descriptions of many applications can be found in the work of Lee [126]. External infrared optical tracking of the wiimote provides the position and orientation of the wiimote at 100Hz update rate and sub-centimeter accuracy.

The features of this approach are:

- Robust tool use: The six degrees-of-freedom pose (position and orientation) of the wiimote are measured at a high update rate (100Hz) with low noise (sub-centimeter). The virtual tools take on the six-degree of freedom pose of the wiimote.
Tool control and gestures: Beyond six degree-of-freedom rigid pose tracking, the
wiimote controls other aspects of the tools through button presses, e.g. turning the
light of the ophthalmoscope on and off, and the virtual hand can form a variety of
poses (e.g. grip, point, lie flat). Button presses are sampled at 100Hz, providing
robust control over the gestures.

Noise-free gesture recognition: With the gesture tool representing the user’s hands,
gestures are chosen at the press of a button. If the user intends to make a gesture
of two fingers held up, he presses the “finger-up” button twice. This instructs the
simulation to display the virtual hand with two fingers held up. The user sees his
virtual hand holding up two fingers and knows that the virtual human also recognizes
that he is holding up two fingers. There is no opportunity for error in the gesture
recognition, in contrast with the other gesture interface approaches described in
Section 5.4.1.

Correct kinesthetic information: The wiimote is manipulated using similar muscle
movements that are used with the real physical tools being simulated.

Passive-haptic feedback: The wiimote provides the weight and shape similar to
many hand-held tools, providing passive-haptic feedback similar to that of the tools
being simulated.

Active-haptic feedback: providing vibratory force-feedback when the virtual tools
collide with the virtual world.

5.4.3 Virtual Hand-Held Tools and Hand Gestures

Three virtual tools were created for MRIPS-CBE: an ophthalmoscope, an eye
chart, and the hand and fingers gesture tool. The user switches between tools by
pressing a button on the wiimote.

5.4.3.1 Ophthalmoscope

An ophthalmoscope is a hand-held tool equipped with a lens for viewing the back
of the inside of the patient’s eye (the fundus) to determine the health of the patient’s
retina and look for symptoms of cranial nerve disorders such as retinal hemorrhages.
The ophthalmoscope moves with six degrees-of-freedom and its position is mapped
one-to-one to the position of the wiimote.
The ophthalmoscope size, shape, and weight are closely approximated by the wiimote, providing passive haptic feedback. As the ophthalmoscope is often held close to the patient’s head and sensitive eyes, it is useful to inform the user if they contact the patient with the (virtually) metal and non-sterile ophthalmoscope. Active-haptic feedback is provided in the form of force feedback when the virtual ophthalmoscope makes contact with the virtual human’s head. If the ophthalmoscope makes contact with the virtual human’s eye, vibratory force feedback is received and the virtual human blinks and jerks his head back.

The ophthalmoscope is typically used to perform two tests: the pupillary reflex test and fundoscopic test. The pupillary reflex test is performed by turning on the light of the virtual ophthalmoscope by pressing the “trigger” button on the rear of the wiimote, and manipulating the wiimote to aim this light into each of the virtual human’s eyes (Figure 5-6 A). The fundoscopic test is performed by turning on the light and moving the ophthalmoscope close to the virtual human’s eye (<8 cm). When this is detected, an image of the fundus of that eye is displayed above the virtual human’s head (Figure 5-6 B). This provides a simplistic simulation of the fundoscopic exam, as use of the ophthalmoscope to view the fundus is simplified, e.g., we use a static image of the fundus instead of an ophthalmoscope-orientation-dependent image. However, our goal is not to train the fundoscopic exam, but to allow a learner to obtain the information provided by a fundoscopic exam for use in diagnosing the cranial nerve disorder.

5.4.3.2 Eye chart

The virtual eye chart is used to test the patient’s visual acuity. The eye chart is fixed in 3D space, much as a physical eye chart is affixed to a wall. Rather than the eye chart position and orientation being controlled by the wiimote, a virtual finger pointing to
a line on the eye chart is manipulated. The eye chart contains eight lines. To point to a line, the user translates the wiimote vertically. The vertical (Y-axis) position of the wiimote is binned into eight intervals corresponding to the eight lines on the eye chart. This illustrates the adaptation of user input to specific tool characteristics. The eye chart is used in the visual acuity test. While the virtual finger is pointing to a line on the eye chart, the user is able to ask the patient to read the pointed-to line, e.g. “can you read this line?” The virtual human reads the currently pointed to line if he is able to depending on the CN affected (Figure 5-7 left). With CN3, 4, and 6, the patient can not read any of the lines with both eyes open. The user is able to ask the patient to cover one eye (Figure 5-7 right) and try again – with one eye closed, the virtual human has 20/20 visual acuity. Alternatively the user can ask the virtual human to report “what is the lowest line you can read?”

5.4.3.3 Hand gesture tool

To provide gesture inputs used in neurological exam tests, a virtual hand and fingers tool is provided. The hand moves with six degrees-of-freedom and its position is mapped one-to-one to the position of the wiimote.

Gestures that can be performed using this tool include making a fist, holding between 1 and 5 fingers up, pointing with one finger, and shaking a finger. The hand tool is used in many of the neurological exam tests. Because of the noise-free gesture recognition, the state of the virtual hand represents both the state of the user’s hand and the hand that the virtual human sees and responds to – there is no ambiguity resulting from gesture recognition error.

**Finger counting test:** The hand can transform from an open hand to a clenched fist. The number of fingers the doctor is holding up can range from 0-5. The up and
down arrows on the directional pad of the wiimote are used to raise and lower fingers. To test the patient’s ability to maintain binocular vision in all fields of gaze, the user will hold one or more fingers up and ask the patient to look straight ahead and report “how many fingers do you see?” (Figure 5-8). If the fingers are held outside of the virtual human’s field of view, he will answer “I can’t see your hand.” For the double vision disorders (CN3, 4, and 6) if the fingers are in view of only one eye the virtual human will report the number of fingers held up; if the fingers are in view of both eyes, the virtual human will report twice the number of fingers held up. If the fingers are in view of only one eye, or one eye is covered, the virtual human will report the correct number. Because of the noise-free gesture recognition, if the virtual human reports twice the number of fingers held up by the virtual hand, the user can be certain that the virtual human is experiencing double vision; there is no ambiguity that the gesture recognition may be malfunctioning.

**Oculomotor (eye movement) test:** By asking the patient to “follow my finger” (alternatively the ophthalmoscope can be used for this test “follow the light”), the user can test the functionality of the patient’s oculomotor muscles and, correspondingly, cranial nerves that innervate these muscles (Figure 5-9). The virtual human holds his head still facing forwards, and attempts to follow the position of the finger with both of his eyes. Because the wiimote and finger move in a one-to-one correspondence, the user receives the same kinesthetic feedback as he would in the real-world exam. This correct kinesthetic information is necessary for learning the psychomotor task of moving the finger in the shape of an uppercase “H” to test the extremes of the patient’s vision.
**Peripheral vision test:** To test the patient’s peripheral vision, the user holds the virtual hand outside of the virtual human’s peripheral vision, instructs the patient “tell me when you see my hand” and then proceeds to move the hand into the patient’s peripheral vision. The virtual human answers “I can see it now” when the finger enters the field of view of either eye. Alternatively, the user can raise one finger on the hand, hold the hand in the patient’s peripheral vision, shake the wiimote, and ask the patient to “tell me when you see my finger shake.” Shaking is detected as changing values in the wiimote’s internal accelerometers.

**Facial sensitivity test:** To test if the patient has feeling in the face, the user can poke the virtual human’s face with one or more fingers and ask “can you feel this” (Figure 5-10). The user knows when he is making contact with the virtual human’s face because vibratory force-feedback is provided by the wiimote when the fingers or hand collide with the face. Collision detection is performed using the meshes of the virtual human’s head and the hand and finger tool, using the OPCODE Optimized Collision Detection library incorporated in the Ogre 3D rendering engine.

It is important to note that if a test requires both user speech and tool manipulation, as is the case for all the tests using the hand gesture tool, the simulation module is designed to perform the test asynchronously or synchronously, whichever is appropriate for the test. For example, “how many fingers do you see?” wants synchronous information, so the virtual human responds (instantly) based on the number of fingers held up on the virtual hand at the time when the simulation module receives the user speech input. However, “tell me when you see my finger shake” and “can you feel this” ambiguously refer to events that could be happening at that point in
time or in the near future, and are handled as asynchronous commands. When one of
these utterances is received by the simulation module, it puts the simulation into a state
in which it is actively looking for a finger shake or a collision between hand and head. If
the finger is shaking, was recently shaking (e.g. within 4 seconds into the past), or
begins to shake within the next 10 seconds (and before another command is given by
the user or the hand tool is deselected), the virtual human will report that he sees the
finger shaking. Allowing for asynchronous events provides more robust communication
– e.g. the user does not have to continuously shake the wiimote while asking the
question multiple times until the two actions coincide.

5.4.3 The Haptic Interface Enhances Communication in Interpersonal Simulation

The tool, gesture, and speech interaction afforded by the haptic interface and
speech interface enhance the communication of the interpersonal simulation. Touch of
the virtual human can be used for communication, similarly to MRIPS-CBE. Touch of
the hand-held tools also enhances communication by providing conversational
grounding.

A common ground is a pool of mutually agreed upon information and serves as a
way to ensure that the message intended to be communicated is received intact by
one’s communication partner [37]. Grounding has been provided in a limited form in
previous interpersonal simulations. For example in Gandalf, a solar system education
application, the user could direct her head gaze to point to a planet and tell Gandalf
“let’s go there” [79]. Grounding is enhanced in MRIPS-NEURO, due to the noise-free
gesture recognition. Because the state of the hand held tools is never ambiguous to
either the human user or the virtual human, the tools serve as grounding objects in
many facets of the neurological exam scenario. Grounding examples (with the
information provided by the tool in parentheses) include: “can you read this line” (which line?), “how many fingers do you see” (held up on the user’s hand), “follow the light” (the light on the ophthalmoscope), and “what’s the lowest one you can read” (lowest line on the eye chart), “can you feel this” (the user’s finger). Just as interpersonal touch enhances communication in MRIPS-CBE, touch for manipulation of hand-held tools enhances communication in MRIPS-NEURO.

5.5 Usability and Content Validity of MRIPS-NEURO for Practicing Diagnoses of Abnormal Findings

The main goal of MRIPS-NEURO is to provide novice learners with increased exposure to abnormal findings in the clinical context of a patient interaction. Learners are able to practice synthesizing physical findings with information gleaned through conversation with the patient. To determine if this goal is met by the design of MRIPS-CBE, we evaluated the usability of MRIPS-NEURO for practicing diagnosing abnormal findings in a focused (on cranial nerves) neurological exam. If a significant proportion of participants are able to correctly diagnose the virtual human using the tests afforded by MRIPS-NEURO, this will also establish the content validity of MRIPS-NEURO.

5.5.1 Study Design and Procedure

An observational study was conducted with nine 2nd-year medical students at the University of Florida. Participants filled out a background survey concerning their experience in neurological examination. All students were considered novices in the neurological exam. However, they had different experience levels with neurological examination:

- All had coursework and some relevant web-based simulator use.
- However, four of the participants had just completed a neurology clerkship (had experience performing neurological exams on human patients) while five of the
participants were just beginning a neurology clerkship and had no experience performing a neurological exam of a human patient.

The distinct difference in experience level is beneficial to the study: more experienced students should be able to reach a correct differential diagnosis (the virtual human had CN3 palsy), while less experienced students may not be able to reach correct diagnosis. If the experienced students can diagnose CN3 disorder in the virtual human, then MRIPS-NEURO is usable for collecting information through speech, gestures, and tool use, and synthesizing this information into a diagnosis in the abnormal neurological exam.

An additional five 2nd and 3rd-year medical students from the Medical College of Georgia were also recruited. All students had completed a neurology clerkship and are considered to be part of the experienced student group. These students completed the same procedure, but a different post-experience survey was used. They are included here only to evaluate whether learners are able to reach a correct diagnosis using the symptoms presented by the virtual human and the tests afforded by MRIPS-NEURO.

Participants completed brief speech recognition volume and quality tests (using Dragon Naturally Speaking 9.5). The participants then began the exam. During the exam, participants could press the “home” button on the wiimote to bring up a tutorial screen for the selected tool (Figure 5-11). After the exam, participants completed a brief survey to assess usability of the interface to complete the exam and report their findings.

5.5.2 Results

All 9 of the more experienced students (4 at UF and 5 at MCG), in addition to 3 of the 5 less experienced students, arrived at the correct diagnosis of CN3 palsy. This
demonstrates the usability of MRIPS-CBE to collect the data needed to arrive at a correct diagnosis in a neurological exam with abnormal physical findings. This also establishes the content validity of MRIPS-NEURO: MRIPS-NEURO simulates the neurological exam scenario to a high-enough degree of fidelity that learners are able to apply their knowledge to arrive at a correct differential diagnosis. Twelve of 14 participants, 85%, a significant majority by a one-way Chi-square test ($X^2(1) = 5.8, p = 0.02$) arrived at a correct diagnosis of a CN3 palsy. Because this is not a large population, the demonstration of content validity should be viewed as a preliminary result. In Section 8.4, a second user study is described, in which 17 of 18 participants arrived at a correct diagnosis of two virtual human patients with a cranial nerve disorder. Over these two studies, 29 of 32 participants were able to use MRIPS-NEURO to gather information regarding symptoms and synthesize this information into a correct diagnosis. This provides strong support of the content validity of MRIPS-NEURO, as 29 of 32 participants is a significant proportion by one-way Chi-square ($X^2(1) = 19.5, p < 0.0001$).

In self-report data of the usability of MRIPS, participants rated the interface and MRIPS-NEURO as usable for performing the neurological exam. In the usability survey, participants rated (1) the usability (dimensions of effectiveness, efficiency, and satisfaction) of the simulator for the physical examination portion of the encounter, (2) the usability of the interface (controlling three tools with one hand-held device), and (3) the usability of each of the tools. The dimensions of usability were rated on the scale: -2(strongly disagree), -1(disagree), 1(agree), 2(strongly agree). Each item was phrased as clearly as possible to avoid confusion. For example, "I was SATISFIED with the
technology during the PHYSICAL EXAM of the patient. I felt comfortable using the technology and did not get frustrated."

Given the small sample, the data for each item was collapsed into a higher level view of usability as "Yes" or "No." This was calculated by summing across each dimension and then using a threshold of 0 (neutral) to determine "Yes" or "No". Results are displayed in Table 5-3.

5.5.3 Observations

Overall, participants were fairly positive concerning the usability of MRIPS-NEURO for performing the neurological exam. Participants were strongly positive concerning the haptic interface. Eight of the nine participants rated the interface positively on all three dimensions of usability. The ophthalmoscope was rated less positively than the other two tools; we expect this is related to issues with tracking, as the physical study configuration did not allow tracking of the wiimote close to the LCD screen on which the virtual human was displayed. Performing the fundoscopic exam tool with the wiimote near the edge of the tracked area, some students tried to move it closer where it could not be tracked.

More experienced participants performed more tests. The less experienced students performed the pupillary reflex test, fundoscopic test, finger counting, eye movement test, and visual acuity test. The more experienced students also conducted the peripheral vision test and other neurological tests unrelated to the eyes (smile, frown, stick out the tongue).

Touch and avoidance of touch was observed. Five of the participants collided the ophthalmoscope with one of the patient’s eyes during the fundoscopic test, and three later avoided another collision when performing the fundoscopic test on the other eye.
One participant was observed colliding with the left eye, then pulling the wiimote back so he could move the ophthalmoscope around the nose, avoiding further contact with the face. During this study, the virtual human did not have the ability to flinch if the eye was contacted by the ophthalmoscope, so all avoidances of touch are due to receiving the force feedback of the wiimote, which appears to be adequate to convey touching.

5.5.4 Conclusions and Continued Evaluation

This observational study provided evidence of the usability of MRIPS-NEURO for its stated goal: provide additional exposure for novice learners in diagnosing abnormal findings in a neurological exam. Additionally, the haptic interface of MRIPS-CBE was rated as usable for performing the neurological exam and appears to be usable for conveying touch. Most importantly, a significant proportion of learners were able to use the hand-held tool use, gestures, and communication affordances of MRIPS-NEURO to correctly diagnose the virtual human’s cranial nerve disorder. This establishes the content validity of MRIPS-NEURO.

The second component of evaluating MRIPS-NEURO is to evaluate the impact of feedback designed to improve learners’ cognitive, psychomotor, and affective performance. This evaluation is presented in Chapter 8. We first provide an overview and motivation of feedback designed for MRIPS-CBE and MRIPS-NEURO.
Table 5-1. List of cranial nerves which can be examined using MRIPS-NEURO.

<table>
<thead>
<tr>
<th>Nerve(s)</th>
<th>Function</th>
<th>Abnormal symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN 2</td>
<td>Visual acuity, visual fields</td>
<td>Poor acuity, peripheral vision loss, abnormal appearance of fundus (retina)</td>
</tr>
<tr>
<td>CN 2, 3</td>
<td>Pupillary response</td>
<td>Pupil does not respond to light</td>
</tr>
<tr>
<td>CN 3, 4, 6</td>
<td>Movement of the eyes, raising of eyelids</td>
<td>Limited movement, ptosis (drooping eyelid)</td>
</tr>
<tr>
<td>CN 5</td>
<td>Facial sensation</td>
<td>Loss of sensation in face</td>
</tr>
<tr>
<td>CN 7</td>
<td>Facial movements</td>
<td>Asymmetry in smile, frown, eyebrow raise</td>
</tr>
<tr>
<td>CN 12</td>
<td>Movement and protrusion of the tongue</td>
<td>Tongue crooked when protruded</td>
</tr>
</tbody>
</table>

Table 5-2. Focused neurological exam tasks and information gained from each to aid in diagnosis of the cranial nerve disorder.

<table>
<thead>
<tr>
<th>Examination task</th>
<th>Information gained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduct medical history interview</td>
<td>Determines what present illnesses, medication, social, family, and sexual history may be involved in the current neurological problem</td>
</tr>
<tr>
<td>Test pupillary reflex</td>
<td>Checks for pupil abnormalities</td>
</tr>
<tr>
<td>Visually examine fundus (rear of inside of eye, e.g. retina)</td>
<td>Checks for intracranial pressure</td>
</tr>
<tr>
<td>Hold fingers up and ask how many fingers the patient sees</td>
<td>Tests patient's binocular vision or double vision</td>
</tr>
<tr>
<td>Move index finger in the shape of an “H” in front of patient’s eyes</td>
<td>Tests for limitation of movement of one or both of the patient’s eyes</td>
</tr>
<tr>
<td>Have the patient read from an eye chart</td>
<td>Tests the patient’s visual acuity</td>
</tr>
<tr>
<td>Move or shake finger in peripheral vision while patient looks straight forward</td>
<td>Tests for peripheral vision disorders</td>
</tr>
<tr>
<td>Ask the patient to blink or wink his eyes</td>
<td>Tests for ptosis, drooping of an eyelid.</td>
</tr>
</tbody>
</table>

Table 5-3. Usability ratings of MRIPS-NEURO (n = 9).

<table>
<thead>
<tr>
<th>Usable?</th>
<th>Participants rating “Yes”</th>
<th>Participants rating “No”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Examination</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Interface</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Ophthalmoscope</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Hand</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Eye Chart</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5-1. An expert performs a neurological exam of Vic, a virtual human patient with double vision due to CN6 palsy.

Figure 5-2. The cardinal eye movements of a normal, unaffected eye. The (yaw, pitch) in degrees of the left eye is displayed next to each depiction.
Figure 5-3. Cardinal movements with the left eye affected by CN3 palsy. In addition to the left eye pointing down and out when the virtual human looks straight, CN3 is notable for ptosis, the drooping of the eyelid of the affected eye.
Figure 5-4. Cardinal movements with the left eye affected by CN6 palsy.
Figure 5-5. The cardinal vectors for normal, CN3, and CN6 eye movements are graphed as (yaw, pitch) pair associated with each axis. The green dot indicates the (yaw, pitch) of the eye when attempting to look straight ahead. The mechanism of the model is visualized for CN6: the vectors $\mathbf{d}$ in the normal and CN6 graph indicate the (yaw, pitch) needed to rotate to the desired gaze position. To find the maximum (yaw, pitch) capable by the CN6 affected eye along vector $\mathbf{d}$, cardinal vectors 1 and 2 are interpolated to calculate the vector $\mathbf{g}$. The magnitude of $\mathbf{d}$ is greater than that of the $\mathbf{g}$, thus the (yaw, pitch) represented by $\mathbf{g}$ is chosen as the final gaze yaw and pitch.
Figure 5-6. A) Testing the pupillary reflex with the ophthalmoscope. B) Performing the fundoscopic test with the ophthalmoscope.

Figure 5-7. Visual acuity test with the virtual eye chart.
Figure 5-8. The finger counting test. A) With both eyes open, the virtual human with a CN3 affected left eye sees double. B) By closing an eye, he sees the correct number of fingers.

Figure 5-9. Checking the eye movement of a virtual human patient with a left eye affected by CN6.
Figure 5-10. Testing facial sensitivity by touching the virtual human’s face.

Figure 5-11. On-screen tutorials for A) the ophthalmoscope and B) gesture tools.
CHAPTER 6
REAL-TIME EVALUATION AND FEEDBACK OF PERFORMANCE

Chapters 6-8 describe the incorporation of real-time evaluation and feedback of user performance into MRIPS-CBE and MRIPS-NEURO. This chapter briefly describes the motivation for feedback; Chapter 7 describes the implementation of the feedback in MRIPS-CBE and Chapter 8 describes the implementation of the feedback in MRIPS-NEURO. A portion of Chapter 7 was published in the proceedings of the IEEE and ACM Symposium on Mixed and Augmented Reality 2009 [127].

Collaborators: Suggestions by medical collaborators D. Scott Lind and Adeline Deladisma motivated creation of the touch map feedback in MRIPS-CBE. One feedback mechanism for affective performance (“thought bubbles”) arose from group discussions with Andrew Raij, Brent Rossen, and Joon Hoa Chuah. In Chapter 8 we briefly describe a previous feedback system for enhancing perspective taking in MRIPS: Virtual Social Perspective-taking (VSP) which was primarily designed by Andrew Raij with contributions from me. Other than the VSP feedback, I designed and implemented all feedback mechanisms.

Relevance to thesis: The thesis states that an interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance improves users’ psychomotor, cognitive, and affective skills in an interpersonal scenario. To prove this statement, we must first develop mechanisms for providing this real-time evaluation and feedback of performance; these are described in Chapter 7 (MRIPS-CBE) and 8 (MRIPS-NEURO). Next, we must demonstrate that the real-time feedback plays a role in improving psychomotor, cognitive, and affective skill sets. Chapter 8 describes a formal study which evaluated the impact of feedback
on these skill sets in MRIPS-NEURO. Evaluation of the impact of feedback in MRIPS-CBE is left to Chapter 9, in which learning and training transfer is also evaluated.

6.1 Motivation for Feedback

Learning takes place as a result of performing a task and reflecting, during and after the task, on how future performance can be improved [12]. However, novice learners are not competent to assess their own performance and thus require feedback from an external source, e.g. expert observer feedback or automated feedback, to initiate reflection [12].

In medical education, it has been shown that immediate, specific, non-judgmental feedback is the single greatest motivator for learning [61][62]. In fact, feedback is necessary from the very beginning of a medical education:

- Students’ treatment of affective aspects of patient encounters (dealing with learner and patient fears and discomfort) become more difficult to modify as learners gain experience.

- Without feedback, learners may develop incorrect skills if these incorrect skills happen to achieve positive outcomes. For example, finding a mass during a clinical breast exam reinforces the technique used by the student even if it is not the correct technique likely to find masses in future exams.

- Additionally, the absence of feedback causes learners to lose desirable behaviors, for example novice medical students instinctively use open-ended questions in medical histories because their knowledge level is low. Without feedback that this process is desirable, students begin to use more close-ended questions (moving to a rigid decision tree) as their knowledge level increases. Close-ended questioning may cause the student to miss a crucial piece of information and reduces patient information disclosure, e.g. about a secondary complaint that may be more urgent than the given reason for coming to see the doctor [12]. Thus feedback is not only beneficial to learning, but the absence of feedback degrades learning and contributes to negative training transfer.

Prior research into the impact of feedback in medical education specifically investigated the feedback given by an expert observer during a physical exam. While expert feedback has the presumed advantage of being of high quality (because it
comes from an expert), this method of feedback has several disadvantages: it is inherently subjective, often presented in a judgmental manner, and provides a high-level qualitative evaluation (“that looks right” or “palpate until you feel the chest wall” vs. more specific and quantitative feedback such as “you have 26% of the breast tissue left to palpate”) [12]. Additionally, this feedback may favor auditory learners who can be “talked through” the exam [42]. These learners make up only 19% of novice medical students [43], leaving the majority of medical students underserved by this approach.

Perhaps the biggest drawback of this approach is the limited opportunities to be observed by an expert – especially in exams in which practice is already limited such as intimate exams and exams with abnormal findings – which leads to many medical students graduating without receiving any feedback concerning their exam performance [4]. When feedback is given, it often comes too late, e.g. at the end of a clerkship or rotation. At this point there are no more opportunities to practice applying what is learned from this feedback. This conundrum breeds residents and practicing clinicians with questionable competence, low confidence, and high anxiety [4][9]. For these reasons, recommendations have been made to provide more detailed (precise), quantitative, objective, and frequent feedback of exam performance [4][9].

Based on these recommendations and evidence of the efficacy of feedback for learning in traditional curricula, we are motivated to incorporate into MRIPS real-time feedback to guide cognitive, psychomotor, and affective skills from real-time quantitative, objective evaluation of performance.

6.2 Unique Capabilities of MRIPS to Evaluate Performance and Provide Feedback

The haptic interface and sensing of its manipulation, in combination with verbal input, provides MRIPS with the novel abilities to quantitatively and autonomously
evaluate and provide feedback concerning learners’ cognitive, psychomotor, and affective skills performance.

As MRIPS is the only interpersonal simulation (as of this writing, we know of no other, published or unpublished) to incorporate verbal and touch inputs, MRIPS is uniquely able to evaluate user performance with data from both verbal and touch channels as well as data from tracking e.g. user head pose and hand position. Previous interpersonal simulations captured only the verbal channel in addition to tracking of user head pose and hand position. These simulations thus had the potential to provide feedback on cognitive tasks involving speech and psychomotor tasks not involving touch. With the addition of the touch input channel, MRIPS is able to evaluate and provide feedback on performance in cognitive, psychomotor, compound e.g. cognitive-psychomotor tasks, and affective tasks.

For example, in the CBE scenario MRIPS is able to provide feedback in the cognitive tasks of taking a breast history and visually inspecting the patient’s breasts; the psychomotor task of palpating with correct pressure; and the compound cognitive-psychomotor task of using a correct pattern-of-search. These psychomotor and cognitive-psychomotor tasks could not be evaluated using prior approaches at interpersonal simulation. Compound cognitive-psychomotor tasks, such as recalling a correct pattern-of-search and recognizing which areas of the breast remain to be palpated can be aided by feedback which incorporates data from the haptic interface’s sensors in addition to tracking data providing the position of the user’s hand. Psychomotor components such as palpating with correct pressure clearly require the haptic interface and its sensing of the force applied to the interface by the user.
Automated evaluation of affective tasks is not as cut-and-dry, as evaluation of affective performance, e.g. determining when proper empathy is used, is largely subjective. However, as touch is a common means of comforting and expressing empathy (both affective tasks) [35], the combination of speech and touch inputs provides MRIPS with more information (than prior speech-only interpersonal simulations) with which to evaluate affective performance. The automated evaluation and feedback of affective performance in MRIPS should be considered a first attempt; the innovation lies in obtaining enough information to make such an attempt.

An example occurs in the clinical breast exam. When the virtual human expresses fear of having a breast exam, it provides an opportunity for the learner to comfort the patient. Touching the patient on the shoulder is a comforting response [35]; MRIPS is able to sense this touch. If the learner makes such a touch in response to the virtual human expressing fear, this touch can be made to elicit a virtual human response expressing her gratitude towards the learner’s comforting gesture. In addition, speech input can also be processed to look for phrases which express understanding (indicating empathy). If the user touches the virtual human’s shoulder while saying “I understand how scary this must be for you with your mother dying of breast cancer,” MRIPS has more confidence that this input should elicit the virtual human to express comfort.

By incorporating a haptic interface augmented with sensing of manipulation of that interface, as well as traditional speech interface and tracking of head pose and hand position, MRIPS is uniquely able to evaluate performance and provide real-time feedback to guide learner performance in the three skill sets.
6.3 Choice of the Visual Channel to Provide Feedback

Feedback has been designed for MRIPS-CBE and MRIPS-NEURO to guide user performance in scenario specific cognitive, psychomotor, and affective skills. All of these feedback elements are presented visually. There are two reasons for choosing to present feedback visually.

We believe that the strength of MRIPS lies in the mixed reality approach – the combination of haptics providing realistic kinesthetic information with the virtual world providing a rich set of visual information – visual information that both recreates the real-world and goes beyond what the real-world can offer. The real world can not provide visual feedback of performance in-situ with physical objects being manipulated, but by registering the real and virtual components of the mixed world, this is easily accomplished. By presenting visual feedback in-situ with the virtual human and her registered (conceptually, but not necessarily physically, co-located) haptic interface, MRIPS avoids additional cognitive load imposed from looking back and forth from a visual feedback presentation and objects being manipulated (as in the approach of Pugh et al [5]; see Figure 1-1). Additionally, the visual channel is hypothesized to have the highest “bandwidth”, explaining the visual dominance (over haptic and auditory senses) [128]. Thus the amount of information that can be presented through feedback, without overwhelming the task at hand, may be maximized by presenting the feedback visually.

In Section 6.1 we described one drawback of expert observational feedback to be its predominantly auditory presentation. While only 19% of novice medical students favor a predominantly auditory learning style, 70% of students favor a learning style combining visual and kinesthetic information presentation [43]. MRIPS already provides
an experience in which visual and kinesthetic information dominate (the exception being the verbal responses of the virtual human), so choosing a visual presentation for feedback serves to further tailor MRIPS to favor the 70% majority of medical students favoring a visual and kinesthetic learning style.

By providing feedback in the visual channel, presented in-situ with the haptics being manipulated and the virtual human being examined, MRIPS can maximize the amount of information provided through feedback and maximize the amount of novice learners who are likely to benefit from this feedback.
CHAPTER 7
FEEDBACK IN MRIPS-CBE

7.1 Introduction

This chapter describes methods of evaluating and providing feedback of user performance in order to guide and motivate improvement in cognitive, psychomotor, and affective components of the clinical breast exam. Each type of feedback is provided in two forms, a real-time visualization presented in the mixed environment and a post-experiential summary of performance presented in a traditional desktop environment.

7.1.1 Cognitive Components

There are two cognitive components of CBE for which automated evaluation and guiding feedback are provided. The first task is to recall the series of questions to ask the patient to take a breast history. The breast history queries the patient’s current complaint (e.g. breast pain) and assesses the patient’s risk of breast cancer. The second task is to recall the poses the patient must assume for the learner to perform a visual inspection of the patient’s breast. These tasks are typically evaluated by expert review of video. In this review, the expert uses a “checklist,” a series of dichotomous (yes/no) items, to rate the learner’s performance. Examples of items on the checklist used to evaluate students in clinic at the Medical College of Georgia are: “Exam: inspected both breasts – arms relaxed” and “[asked about] Risk Factor/Symptom: Breast self-exam.” In MRIPS, feedback is provided to guide learners in performing complete breast histories and visual inspections. The feedback is provided in the form of a list of questions to ask (text) and visual inspection poses (icons). This procedural checklist (Section 7.2) changes appearance in real-time as the learner progresses through the breast history and visual inspection. An item is automatically “checked-off”
the list when the learner asks a question related to the item. This relies solely on the speech interface which, while present in prior interpersonal simulations, has not been used to provide real-time feedback for similar cognitive tasks. This type of feedback could be provided with SPs, though it would require additional effort by the SP to provide this feedback during the interview and exam, and thus is not typically provided in real-time [4][16].

7.1.2 Affective Components

Affective components of CBE are to alleviate the patient’s anxiety and address the patient’s concerns – effectively maintaining the patient’s comfort level – by engaging in perspective taking and expressing empathy (understanding of the patient's concerns) when appropriate. Other than patient speech, the patient’s facial expressions and tone of voice are the only cues to guide learners in these tasks. Because novice learners are under a high cognitive load during a CBE, they perform poorly in recognizing these cues and do not engage in affective elements of the exam [39]. MRIPS-CBE provides learners with an additional cue to guide learners to consider the patient’s concerns and emotional state. These cues take the form of “thought bubble” feedback (Section 7.3), in which the virtual human’s internal state (emotions, concerns, feelings towards the learner) are presented in the visual form of a cartoon thought bubble. Thought bubbles are used to prompt learners when empathy or comforting is appropriate (cuing learners’ empathic responses) and to show learners how their speech and actions can negatively or positively change the virtual human’s affective state (providing feedback on appropriateness of learners’ empathic or non-empathic responses).
7.1.3 Psychomotor Components

The psychomotor components of CBE are palpation of the breast using the middle three fingers. Palpation should be performed using the middle three fingers (finger pads) pressing in overlapping circular motions at three successive levels of pressure. Each palpation covers these three levels in sequence: light (subcutaneous), medium (midlevel), and deep (to the chest wall) [9]. This is the recommended procedure known as the Mammacare method. We chose to evaluate users of MRIPS-CBE according to this method because it is known to outperform other methods at finding breast masses [66], and it is the method taught at both the Medical College of Georgia and the University of Florida College of Medicine, with whom we collaborate.

The sensing approach used in MRIPS-CBE is not able to accurately distinguish palpation with three fingers from palpation with e.g. two fingers, so evaluation and feedback is not provided for this aspect of palpation. Collaborating medical experts did not consider this a concern because remembering to use three fingers is considered minor compared to palpating with correct pressure. Feedback of the use of pressure is provided in the form of the touch map visualization, a series of color coded visual elements overlaid on the virtual human’s breast (Section 7.4).

Feedback is also provided for two compound cognitive-psychomotor tasks. The first is recalling and following the correct pattern-of-search in which to palpate. Learners’ patterns-of-search are guided by the pattern-of-search map visualization, which displays an expert’s pattern and the learner’s deviation from the expert pattern in real-time (Section 7.5). The second compound cognitive-psychomotor task is recognizing which areas of the breast remain to be palpated (coverage of the breast). This task is guided as part of the touch map visualization described in Section 7.4.
7.2 Procedural Checklist

To guide learners through the cognitive tasks of 1) recalling what questions to ask in order to evaluate the virtual human’s breast cancer risk and 2) recalling the poses for visual inspection, we provide a graphical checklist of text and icons. An item on this checklist is automatically “checked off” or highlighted when a user utterance is matched to a virtual human response corresponding to the item. The checklist items are divided into three parts: breast history (Figure 7-1), visual inspection (Figure 7-2), and palpation (Figure 7-3). To illustrate how this visualization changes to guide participants through the cognitive tasks, in Figure 7-1 part B, the learner has asked the virtual human about the location of the pain in her breast and whether she has experienced nipple discharge or any other changes in her breasts.

Novice medical students are under a high cognitive load during the CBE. This high cognitive load causes novice learners to miss elements of cognitive tasks as these require verbal communication; learners are too focused on performing a correct manual exam (palpation) to communicate effectively [38].

It is the goal of the breast history procedural checklist feedback to help to reduce this load by guiding students through the task of assessing the patient’s cancer risk. The checklist includes important questions to ask divided by topic (history of present illness, medical history, family history, social history). This feedback may also help to keep students on topic, as Raij et al. previously found that novice medical students have a difficult time following a logical sequence of topics but instead jump from topic to topic. This behavior causes them to miss important questions which may lead to failure in a CBE of a human patient [84]. The breast history procedural checklist items are taken
from validated instruments used to evaluate medical students at the University of Florida’s College of Medicine and at the Medical College of Georgia.

The palpation checklist in Figure 7-3 is designed to primarily aid the cognitive task of recalling the areas of lymph nodes peripheral to the cone of the breast which also require palpation. The palpation procedural checklist changes appearance in response to the learner’s palpation of these areas on the MRIPS physical interface.

The completeness of a learner’s CBE with respect to breast history, visual inspection, and palpation of peripheral breast tissues is also provided after the MRIPS-CBE interaction. This feedback is provided as part of a desktop-based post-experiential feedback application (Section 7.6). The overall goal of the real-time and post-experiential procedural checklist feedback is to guide novice learners through performing a more complete CBE.

7.3 Thought Bubbles

The thought bubbles feedback augments the communication between virtual human and human to include visual representations of the virtual human’s thoughts. This feedback has the goal of informing the user how well he is handling the affective components of the CBE. Through thought bubbles, the virtual human indicates when she requires comforting (is scared, anxious, sad). The user’s attempts to comfort the patient through expressions of empathy and comforting touches are evaluated by capturing the user’s speech and touching of the haptic interface. In response to a comforting attempt by the user, the virtual human indicates the effect that the user’s actions have on her emotional state through thought bubbles and speech.

Expressing the virtual human’s emotional state through thoughts demonstrates to the learner that his patient may be emotionally affected by the learner’s actions and
words – even if the patient does not verbalize these emotions. Prior work in medical education has shown that novice medical students are largely unaware of the patient’s emotions [129]. Showing learners how their actions change the patient’s emotional state may help prime learners to be more aware of their patients’ emotions.

Previous work in visualizing internal state of virtual humans in interpersonal simulations has been limited to the ELECT BiLAT system [83], a negotiation and cultural competency trainer. In ELECT BiLAT, the emotional state is depicted as text hovering above the virtual human’s head. A later version included other internal state information visualized in the style of progress-bars, e.g. a green line progressing between 0% trust and 100% trust of the user.

We believe that our visualization of emotional state as thoughts allows for more realistic and detailed emotional expressions than these prior attempts at visualizing virtual human emotions. As an example, the virtual human can think “the doctor seems more concerned with the exam than with my feelings” instead of text indicating that she is “annoyed” or a graph indicating 10% sadness.

Thought bubble feedback is used to indicate when the virtual human is in need of comforting, when she has been comforted, and when the user has failed to comfort her. Comforting of the virtual human through speech and touch can only be triggered when the virtual human has first prompted for comfort, putting the simulation in a state in which it recognizes speech and touch inputs as comforting or non-comforting. Thus, we are able to quantitatively evaluate affective performance as the percent of opportunities for comforting in which the learner successfully comforts the patient. The goal of this feedback is to aid novice learners in becoming cognizant of emotional situations within
interpersonal communication and addressing these situations with appropriate expressions of empathy.

7.3.1 Automated Evaluation of Affective Performance

Determining the amount of empathy expressed by a user’s response is highly subjective, however in our experience with nearly 100 prior users of MRIPS-CBE, there is a clear distinction between responses that empathize, sympathize, patronize, or ignore emotional content (e.g. by moving on to a new topic). Empathic responses are distinct in their expression of understanding, while sympathetic responses express being “sorry,” and patronizing responses take the form of an instruction, e.g. “don’t be scared” or “there’s nothing to be scared about.” Responses may also ignore the emotional content of the patient’s speech. For example, when the virtual human states “I lost my mom to breast cancer two years ago; I miss her everyday,” examples of empathic, sympathetic, patronizing, and ignoring responses that have been elicited in user studies of MRIPS-CBE are:

- **Empathic:** “That must make this [exam] very difficult for you.” “I understand how difficult it can be to lose a loved one.” “Are you handling it ok? Have you been able to talk to anyone about your feelings?”
- **Sympathetic:** “I’m sorry about your mother.” “I’m sorry to hear that.”
- **Patronizing:** “Don’t be sad. It’ll be ok.”
- **Ignoring emotion:** “How old was your mother when she died?”

To determine if a user utterance was empathic, sympathetic, or patronizing, we grouped user responses given in prior evaluations of MRIPS-CBE (Chapter 4) which were rated by experts for their empathic content on a scale of 1-5, where ratings of 3-5 indicated empathy, and for their appropriateness (also from 1-5, 1 being extremely inappropriate, 3 being neutral, and 5 being an exemplary response). Experts also rated responses as empathic, sympathetic, or both. Reducing the appropriateness rating to a
dichotomous (inappropriate / appropriate) variable tended to group empathic and sympathetic responses together as appropriate and to group patronizing and ignoring responses together as inappropriate. We then categorized other non-rated responses by their similarity, albeit subjective similarity, to the expert-rated responses and extracted key phrases for each of the groups.

However, instead of trying to differentiate between these four categories – as the expert ratings did not specifically do this – we chose to group utterances based on the expected effect on the patient’s comfort level and affect (positive or negative) towards her doctor. Thus user utterances could trigger a virtual human expression of being comforted, an expression of not being comforted but recognizing the user’s attempt to comfort her, and an expression of negative affect towards the user.

User utterances which contained empathic key phrases triggered the virtual human to indicate that she was comforted by the user’s speech. Additionally, we allowed touching of the shoulder or arm of the haptic interface to comfort the virtual human, as we observed touching of the virtual human’s upper arm to comfort the patient in prior observational studies [82][93] and touch is the most commonly used method of comforting [35]. This response is aimed to give the user reinforcement of appropriate handling of the emotional situation.

Utterances which contain key phrases indicating a sympathetic or patronizing response triggered the virtual human to indicate that she was not comforted by the user’s speech but understood that the user was trying to comfort her. This response is aimed to indicate to the user that he needs to handle the emotional situation with a more empathic approach.
If the user responds to the virtual human’s prompt for comforting by ignoring the emotional content, e.g. moving on to another topic which may or may not be related to the prompt for comforting, the virtual human’s response indicates negative affect towards the user, because the user is not paying attention to her feelings.

This is certainly a limited approach to evaluating empathy, and should be considered only a first attempt which produces only an approximate rating of empathic performance. In addition to touch, this attempt considers only the words spoken by the user, not the prosody (inflection, rhythm, and stress) which is an important conveyer of emotional content [130]. Additionally, the reliance on speech recognition to provide these words will cause some attempts at empathy to go unrecognized, as speech recognition performance is less than perfect. In the future we hope to augment this detection of empathy with a battery of further sensing, e.g. user posture and prosody.

### 7.3.2 Feedback to Reinforce and Correct Affective Performance

The feedback elements take the appearance of cartoon thought bubbles, as this is a common (in American culture) method of expressing when a person is thinking. The thought bubble and its two trailing bubbles are textured quadrilaterals that always face the camera (i.e. billboards). A simple heuristic is used to ensure the visibility of the bubble. The bubble has an affinity for the right side of the patient’s head (from the user’s point of view), as the left side is occupied by the procedural checklist. However if there is more screen space above or to the left of the virtual human’s head, the bubbles will appear in that location instead. When the patient is lying down for palpation, the bubbles appear above her head.
Two different situations requiring comforting (empathic responses) and the virtual human responses to correct and incorrect handling of these situations are depicted in Figures 7-4 and 7-5.

### 7.4 Touch Map

The touch map evaluates the use of correct palpation pressure at light, medium, and deep levels and presents feedback to reinforce correct use of pressure and indicate the need for correction of incorrect (too-hard) pressure. This is accomplished by comparing the learner’s use of pressure at each palpation to the pressure used by an expert in a pre-recorded CBE. Thus, the touch map is able to evaluate the novice learner’s use of correct pressure in relation to an expert’s use of correct pressure. The touch map also guides and evaluates palpation of the entire breast. Completeness is also evaluated in relation to an expert’s CBE. The process of providing feedback of palpation pressure and palpation coverage completeness is to first capture the palpation pressures and positions of an expert’s CBE (Section 7.4.2), processing this data to define light, medium, and deep pressure levels (Section 7.4.3), and then, during the learner’s exam, determining in what pressure level the learner’s palpations belong (Section 7.4.4).

#### 7.4.1 Feedback Goals

The goal of the touch map feedback is to provide a precise and quantitative evaluation of palpation pressure to equip learners with a skill that will lead to more complete CBEs and more effective detection of breast masses. Feedback of palpation pressure with the precision of touch maps has not previously been available. The correctness of a learner’s palpation pressure can not be evaluated without the sensing approach taken by MRIPS and the simulator of Pugh et al. (Figure 1-1) [39]. Touch
maps expand on the feedback provided by Pugh et al. MRIPS uses more sensors in a more dense configuration than Pugh et al., evaluates all three levels of pressure instead of two (hard, too-hard), and provides continuous feedback of pressure over the entire breast instead of at discrete areas on the breast.

### 7.4.2 Capturing Palpation Pressure and Pressure in an Expert’s CBE

In order to determine what constitutes correct light, medium, and deep pressure and complete coverage, an expert performs an exam using MRIPS-CBE. Palpation position and pressure data captured during this exam is later processed to model the three pressure levels (and a too-hard pressure level), correct pattern-of-search (Section 7.5), and complete coverage. An expert exam is performed once per setup of the MRIPS system, as during setup, the image-space of the augmentation and infrared-tracking cameras is registered with the physical interface. Any number of learner exams can be performed after this calibration step; in our installation at the Medical College of Georgia, over 50 exams have been performed by students without need for recalibration.

The expert’s palpation pressure is defined as the set of 64 floating point values reported by the set of 64 force sensors embedded in the haptic interface of the virtual human (detailed in Chapter 3). The force sensors are sampled at 35 Hz with a maximum delay of ~90 ms between applying a force to the haptic interface and receiving the set of 64 sensor values. The values received by the MRIPS simulation module are relative to automatically made baseline measurements. The rest of the process of displaying feedback takes ~30 ms, resulting in a total of slightly more than one tenth of a second of delay between palpating and receiving feedback. This upper bound on delay was measured by forcing the application to block (wait) on a new
camera frame and set of sensor data, and measuring the time from receiving both to displaying an updated frame. This delay is acceptable as the MammaCare method of CBE calls for each palpation motion to have a duration of 1-2 seconds.

Palpation position is calculated in the image space of the color camera that provides the MRIPS video augmentation. The touch map feedback is later rendered into this video stream. This video stream is in turn projected onto the mesh of the virtual human, using a projected texture. This process displays the touch map in-situ with the physical breast being palpated. The expert wears a piece of infrared reflective tape on the fingernails of her middle three fingers. The positions of all pixels belonging to this piece of tape are captured by the infrared camera paired with the color augmentation camera (the pairing is again shown in Figure 7-7). The transform from a pixel in the infrared camera to a pixel in the color camera has previously been calculated. This transform is simplified to three degrees of translation and one degree of rotation, and is calculated by waving an infrared marker around in view of both cameras. The marker position is found in each camera’s image space, and the transform is calculated by applying Horn’s algorithm [131] to the resulting point clouds. After transformation into the color camera’s image space, all pixels belonging to the infrared marker are recorded, along with their centroid and a logical timestamp (an integer number used to order the sets of pixels temporally).

Throughout the expert exam, the area of the infrared marker is estimated. The mean estimate is taken to be the area of a circle defining the size of a palpation motion. The radius of this circle is stored and later used to draw the feedback elements during the learner’s exam.
7.4.3 Guiding and Evaluating Complete Coverage

The union of all infrared marker pixels from each camera frame of the expert’s exam represents the area that must be palpated for a complete exam (complete coverage).

The boundary of this area is presented to the learner as part of the touch map feedback to indicate the area needed for a complete exam (Figure 7-8). To find this boundary, simple edge detection is first performed. Minor smoothing of this boundary is needed because of noise in the detection of the infrared pixels. This noise can produce “cracks,” thin areas in which the expert palpated but infrared pixels were not detected. These are undesirable – while the expert area need not be convex, the shape should be fairly simple and there should not be gaps in the area within the cone of the breast.

To achieve these desired qualities, the boundary is smoothed. First, boundary pixels are ordered by walking the boundary clockwise and discarding pixels that are farther than five pixels from the current ordered boundary (the distance of 5 pixels is governed by the earlier edge detection method). This removes sharp changes of direction of the boundary line. The ordered boundary pixels are then filtered, to smooth the appearance of the boundary line, by convolving with the 1D filter \(\{\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\}\).

To calculate the area bounded by this smoothed boundary line, the frame is flood filled starting from a pixel known to be outside the boundary. The inverse of the filled area now represents the area required for a complete exam. The number of pixels in this area will be used to calculate the percentage of breast tissue the learner has palpated. The frame created from this process (pixels with alpha = 1.0 represent the boundary; 0.5 inside the boundary; and 0.0 outside the boundary) is retained to use as a mask for later calculations of the area the learner has palpated.
During the learner’s exam, each palpation creates a circular element whose color (and alpha value) represent the pressure level. These elements are accumulated in a texture functioning as an accumulation buffer. Parts of elements that extend beyond the boundary are discarded using the boundary mark. Thus at any time during the learner’s exam, the total area that has been palpated by the learner is simply the number of pixels with alpha value greater than zero in the accumulation buffer. The buffer’s pixels are also processed to count the number of pixels belonging to each of the light, medium, deep, and too-hard pressure levels. This produces a measure of the area of the breast palpated by the learner, in pixels. The percentage of breast tissue that the learner has palpated at each pressure level (and total percentage) can then be calculated by dividing by the previously calculated area (in pixels) required for complete coverage. These percentages are provided to the learner in the post-experiential feedback (Section 7.6; Figure 7-15).

As the learner palpates, he sees the area of the breast within the boundary becoming covered with touch map elements, guiding the cognitive task of palpating all breast tissue.

7.4.4 Calculating the Palpation Pressure Levels

After capturing the palpation pressure and position for the expert’s complete exam, this data is processed to create a model of correct pressure at the three required pressure levels (light, medium, and deep) and an inappropriate “too-hard” level of pressure.

Modeling correct pressure consists of determining the range of sensor values corresponding to each pressure level at each possible palpation position. Because of the high dimensionality of the sensor data (64 real values, one from each sensor) and
the large size of the space of possible palpation positions (order of $10^5$ in the 640x480 image), we instead model the pressure ranges at each sensor. During the learner’s exam, the pressure level is calculated at each sensor, and then the pressure level at the palpation position is calculated as a weighted average of the pressure levels at the sensors.

Modeling pressure ranges at each sensor avoids the computational expense of working with high dimensional sensor data. At each sensor, a pressure level can be modeled in one dimension— if pressure levels were modeled at each palpation position, one dimension per sensor would be required. This approach also provides the option of using the sensor values to estimate palpation position. However, in evaluations we have calculated learner palpation position by tracking a 0.25 cm radius infrared marker on the learner’s middle fingernail.

Modeling pressure levels at each sensor: The low, medium, and high pressure ranges are naturally present in the sensor data of the expert’s exam. Calculating these ranges is an unsupervised learning problem which can be solved using clustering. A Gaussian mixture model (GMM) with three one-dimensional components (corresponding to light, medium, and deep pressure levels) is fit to the set of non-zero values reported by each sensor during the expert’s calibration exam. Each component of the GMM takes the form of Equation 1-1.

\[ N_{GMM,k} (v_s | \mu_k, \sigma^2_k) \]  

(1-1)

Initial values for the GMM’s expectation-maximization algorithm are provided by first applying $k$-means ($k = 3$) to the data. The too-high pressure level is modeled as an
additional one-dimensional Gaussian component, which is a shifted (an experimentally
determined shift of +2.5 std. deviations) duplicate of the high pressure component.

Modeling the influence of each sensor: The relationship between the values
reported by a sensor \( s \) and the possible palpation positions are modeled as a 2D
Gaussian centered at the position of sensor \( s \) in image-space (the mean of the 2D
Gaussian). The position of sensor \( s \) is estimated as the weighted mean of expert
palpation positions corresponding to non-zero values of sensor \( s \), using the values of
sensor \( s \) as the weights. To reduce the impact of noise in the sensor data, this
calculation includes only those palpation positions corresponding to values of sensor \( s \)
that are one std. deviation greater than the mean value reported by sensor \( s \) during the
expert exam. This adaptive thresholding heuristic calculates the sensor’s position in
image-space to within the radius of the sensing element, resulting in \( \sim 0.5 \text{cm} \) (or \( \sim 5 \)
 pixels) of error. The covariance of the 2D Gaussian is calculated as the weighted
covariance, again with the sensor’s values as weights, but without thresholding. After
using the expert’s exam data to calculate the 2D Gaussian of Equation 1-2 for each
sensor, the learner’s palpation position can be estimated by Equation 1-3, where \( v_s \) is
the value of sensor \( s \).

\[
N_{2,s} (\bar{x}_{\text{img}} | \mu_{2,s}, \Sigma_{2,s})
\]

\[
\bar{x}_{\text{img}} = \frac{\sum \mu_{2,s} * v_s}{\sum v_s}
\]

\[
l_s = \frac{\sum k * N_{\text{GMM},k} (v_s | \mu_k, \sigma_k^2)}{\sum k N_{\text{GMM},k} (v_s | \mu_k, \sigma_k^2)}
\]
Calculating level of pressure at the learner’s palpation position: The model of correct pressure can be described completely by the four-component 1D GMM and 2D Gaussian at each sensor. During the learner’s exam, the model is evaluated at the set of sensor values $V$, and the model returns a continuous value in the range $[1 = \text{light}, 4 = \text{too-high}]$ representing the learner’s palpation level. Given the set of sensor values $V$ and the learner’s palpation position $x_{\text{img}}$ reported by 2D infrared tracking or calculated using Equation 1-3, the learner’s palpation pressure level is calculated using Equation 1-4 and Equation 1-5.

For each sensor $s$ with non-zero value $v_s$, calculate the pressure level $l_s$ at sensor $s$ using Equation 1-4. The pressure level $l_x$ at $x_{\text{img}}$ is then calculated by Equation 1-5 by interpolating between pressure levels at each sensor.

The correctness of this model is evaluated informally by demonstrating that it produces, for the range of all sensor values, output consistent with 4 distinct levels of pressure. It is expected that as the user reaches either of the light, medium, and deep levels of pressure, there is a range of sensor values for which the pressure remains in the same level. The values of these ranges vary with the thickness of the breast tissue and cannot be known a priori – these ranges are discovered by fitting of the GMM to the expert calibration data. Also, as the user transitions between levels, the output of the model should be approximately linear, as the value returned by a force sensor scales linearly with the force applied. The model reproduces this expected behavior, as shown in Figure 7-9.
We have explored other methods for modeling correct palpation pressure, including the naïve approach of finding a nearest neighbor in the expert’s sensor data to the learner’s sensor data, but the high-dimensionality of the sensor data makes these approaches too computationally expensive to evaluate in real-time. In contrast, the presented model is computationally inexpensive (evaluation of 5 Gaussian distributions at each active sensor with typically no more than 5 sensors active at once), allowing the learner’s use of correct pressure to be evaluated in real-time, to guide, reinforce, and correct the learner’s palpation pressure.

7.4.5 Design of Feedback Elements to Guide, Reinforce, and Correct

The touch map provides visual feedback of the learner’s use of correct pressure and coverage of the breast. The touch map applies two rules to present this information visually: the coverage is encoded as the feedback element’s shape, a circle, and the pressure is encoded as the color of the feedback element, a multicolored scale with distinct colors at each of the four pressure levels.

Because each palpation consists of applying pressure using a circular motion, the shape of each visual element of the touch map is a circle. The radius of the circle is calculated during the expert calibration exam to provide a circle of approximately the area that the expert’s middle finger covers during the palpation motion. Each palpation of the breast results in one of these circular elements. The union of these circles represents the area of the breast tissue palpated, guiding the learner in palpating the entire breast.

The level of pressure (light, medium, deep, too-high) the learner uses is represented by the color of this circle. A multicolored scale with a distinct color at each pressure level was chosen, as multicolored scales are preferred for identification tasks.
(i.e. identifying areas of the breast which have not been palpated with light, medium and deep pressure) [132]. The colors chosen for each pressure level are influenced by prior information visualization literature and discussion with medical educators. The ability of the color scale to convey use of correct pressure was informally evaluated by feedback from medical students (Section 7.7). As a blue-green-yellow scale best encodes continuous data [133], these colors are chosen for the low, medium, and high pressure levels (low = blue, medium = yellow, high = green). The order of green and yellow were swapped so that green’s connotation with “good” would match “good” to reaching the high pressure level. Red was chosen for the too-high pressure level, as red connotes “stop.”

Given the continuous pressure level value $l_x$ outputted by the model of correct pressure, the color of the visual element is calculated by linearly interpolating between the colors at the neighboring pressure levels floor($l_x$) and floor($l_x$)+1.

The color of the element guides use of correct palpation pressure and indicates the need for correction of incorrect palpation pressure (Figure 7-10). The learner is guided to increase pressure until the element becomes colored green. Reinforcement of correct pressure occurs by seeing the element colored green, associating the learner’s muscle movements with the knowledge that correct pressure was applied. As the pressure applied begins to exceed appropriate deep pressure, the element begins to turn red, indicating to the learner to stop increasing pressure. Correction is provided from seeing the element colored red, indicating that inappropriately high pressure was used.
7.4.6 Presenting Feedback In-situ with the Virtual Human and Physical Breast

An improvement of MRIPS over previous approaches to providing feedback of palpation in intimate exams (e.g. Figure 1-1) is to display the feedback elements in-situ with the anatomy being palpated. The learner needs only to look in one place, which will reduce the increase in cognitive load imposed by the feedback. To achieve in-situ display of the feedback elements, we render the feedback elements into a real-time video stream of the physical breast.

An image-based approach is taken to locating the visual feedback in-situ with the virtual human and the physical breast of the haptic interface. The visual elements of the touch map and the pattern-of-search map are rendered into the real-time video stream of the learner’s hands and the physical breast, which augments the virtual human. Fragment shader programs and render-to-texture are used to simulate an accumulation buffer for each of the touch map and pattern-of-search map. As the touch map accumulates visual elements, existing color is overwritten only by color representing higher pressure. The final color of each touch map element thus represents the highest pressure used at that position. For the pattern-of-search map, more recent elements occlude older elements. For each frame of video, the latest touch map and pattern-of-search map, in that order, are alpha blended with the video frame. The video frame is then projected onto the mesh of the virtual human using a projected texture. The result is that the feedback appears located on the virtual human’s breast.

7.4.7 Design Choices

7.4.7.1 How many experts are needed to model psychomotor performance?

We chose to base the model of expert palpation pressure on a single expert’s CBE. Using a single expert is standard in mixed environments which seek to teach
psychomotor skills through mimicking a pre-recorded expert performance; e.g. learning
tai-chi [134], tennis [135], rehabilitation [136], and laparoscopy [137]. However, the
validity of the model of expert performance may be increased by incorporating data from
multiple experts. The model of palpation pressure is trivially extended to multiple
experts; multiple expert calibration exams are performed and the aggregate data
processed. However, every expert will use slightly different amounts of pressure, so
with multiple sets of expert data we expect that it may require many experts to find
distinct clusters representing the three pressure levels. In this respect, the approach of
using a single expert may prove to be more practical, but with the risk of being less
representative of a large population of breast examination experts.

7.4.7.2 Visual feedback elements occlude the learner's hands

The touch map and pattern-of-search map are rendered “on-top” of the video
stream of the learner’s hands and the physical breast, with a learner adjustable partial
transparency. This maximizes the visibility of the visual feedback. However this may
make it difficult for learners to locate their fingertips in the virtual world. We have
experimented with adaptive-k Gaussian mixture models [138] for segmenting the
learner’s hands from the video stream, in order to render the feedback under the hands,
but this has not been incorporated in MRIPS-CBE because of the high computational
complexity.

7.4.7.3 Drawbacks of an image-based approach

Because we take an image-based approach to locating the visual feedback in-situ
with the virtual human and physical breast, a new expert calibration exam is required if
the cameras are moved with respect to the physical breast model. However, a one-time
installation of MRIPS-CBE at the Medical College of Georgia has been used in more than 50 interactions without the need to recalibrate.

7.5 Pattern-of-Search Map

The pattern-of-search map evaluates and provides feedback of the learner’s following of an expert’s correct pattern-of-search, guiding, reinforcing, and correcting the learner in the cognitive task of recalling and following a correct pattern-of-search.

7.5.1 Feedback Goals

Following a systematic pattern of search, such as the vertical strip pattern, is known to result in CBEs which are more successful in finding breast abnormalities [4][9]. The goals of the pattern of search feedback are to allow a medical educator to define a pattern of search for learners to follow; and to guide learners to follow this pattern to develop more successful CBE technique.

7.5.2 Modeling Correct Pattern-of-Search

A model of correct pattern-of-search takes as input a recent subset of the learner’s palpation positions, and outputs a quantitative measure of the learner’s deviation from expert pattern-of-search. Modeling correct pattern-of-search consists of recovering the expert’s pattern from the palpation position data of the expert calibration exam, and creating a real-time evaluable model to calculate the learner’s deviation from this expert pattern.

\[
P = \{\vec{p}_1, \ldots, \vec{p}_n\}
\] (1-6)

Recovering the expert’s pattern-of-search: The set of palpation positions captured in the expert calibration exam is given by Equation 1-6. This set contains clusters corresponding to each distinct palpation. This is shown for a vertical strip pattern in
Figure 7-12A. The centroids of these clusters are calculated by processing $P$ in temporal order and creating a new cluster when the distance between the current cluster’s centroid and the next position $p_i$ is greater than the radius $r$ of a circle representing the area the expert’s fingertips cover in each palpation. Resulting centroids are shown in Figure 7-12B. Because the noise present in the IR tracking of the expert’s palpation positions influences the centroid calculation, the centroids are then filtered in temporal order by convolving with the neighborhood $(\frac{1}{4}, 1, \frac{1}{4})$. The final expert path is created by constructing a Catmull-Rom spline with the filtered centroids as control points (Figure 7-12C). The Catmull-Rom spline was chosen because it passes through all control points. Direction indicators are added when rendering the expert path (Figure 7-12E). The spline reconstruction of the expert pattern is stored as a sequence of line segments, $S$, which will be used to evaluate the learner’s pattern – also represented a sequence of line segments $L$ between the learner’s successive distinct palpation positions.

7.5.3 Guiding and Evaluating Learner Pattern-of-Search

In CBE, the learner’s goal is to use the same pattern-of-search as an expert, but it is not necessary to follow the expert’s trajectory exactly – for example, the learner should not be penalized for following the pattern in reverse temporal order or for a small translation between learner and expert patterns. Thus, Euclidean distance is a poor metric of learner deviation from the expert pattern. Deviation from the expert pattern is instead taken to be the angle between matched segments of learner and expert patterns.

We experimented with a naïve approach to matching learner and expert patterns: the two nearest (in Euclidean distance) expert segments to the current learner segment
were found, and the deviation calculated as the average of the angles between the learner segment and the two expert segments. However, this approach exhibited poor performance in portions of the pattern with high curvature and penalizes small translational offsets between learner and expert.

An approach which avoids these problems is to not explicitly define a matching between learner and expert segments, but instead to create from the expert pattern a vector field which serves as a look-up-table.

Our approach is to place radial basis functions, of the form of Equation 1-7, at the midpoints of the line segments of the expert’s pattern, where \( m_i \) is the midpoint of segment \( s_i \) and \( r \) is the radius of the circle representing the area of each palpation. Each radial basis is associated with the normalized vector \( \hat{s}_i \) of the line segment at which it is placed. The vector field value at \( x_{\text{img}} \) is calculated by Equation 1-8. Instead of storing the vector field, a scalar field is stored to simplify computation during the learner’s exam. The scalar field contains the absolute value of the dot product of \( v(x_{\text{img}}) \) with the reference vector \((0,1)\). The absolute value causes forward and reverse traversal of the pattern to be equivalent. This scalar field is visualized in Figure 7-12D.

\[
f_i(x_{\text{img}}) = \exp(-r^{-2} \|x_{\text{img}} - \hat{m}_i\|^2) \quad (1-7)
\]

\[
\bar{v}(x_{\text{img}}) = \sum_i \left\{f_i(x_{\text{img}}) \ast \hat{s}_i\right\} / \sum_i f_i(x_{\text{img}}) \quad (1-8)
\]

To calculate the deviation of the current line segment of the learner’s pattern, the scalar field values \( s(x_1) \), \( s(x_2) \) at the endpoints of the segment are retrieved and the dot product \( d \) between the learner’s current line segment line and the reference vector calculated. The learner’s deviation from the expert pattern is then calculated as the average of the differences \(|d-s(x_1)| \) and \(|d-s(x_2)|\).
This method affords quantitative feedback of the learner’s performance relative to the expert: the number of learner palpations vs. the number of expert palpations; the percentage of learner palpations which correctly followed the expert pattern (deviation of < 15 degrees); the percentage of learner palpations which fell into low deviation (< 15 degrees), medium deviation (15-30 degrees), and high deviation (>30 degrees); and the total amount of deviation of the learner’s pattern from the expert’s pattern. The number of learner and expert palpations and the percentage of learner palpations correctly following the pattern were chosen by our medical collaborators as the most meaningful to learners; this numerical feedback is incorporated into the post-experiential feedback described in Section 7.6 and depicted in Figure 7-15.

7.5.4 Design of the Feedback Elements for Guiding, Reinforcement, and Correction

The pattern-of-search map encodes the progression of the search pattern as a series of arrows, and encodes the deviation of the student’s pattern from the expert’s pattern as a multicolored scale.

Patterns of search are typically presented in medical texts as a series of arrows (e.g. [9]). Thus, the series of line segments which reconstruct the learner’s pattern are visualized as a series of arrows which point towards increasing time. The appearance of an arrow is constructed in real-time as a series of polygons, rather than offline-created sprites. The rendering of each arrow is parameterized with parameters of arrow tail and head widths, arrow length, and arrow color.

The color of each arrow represents its deviation from the expert’s pattern-of-search. We chose a three-colored scale of green, yellow, and red, as with traffic lights (go, caution, stop). Green encodes that the student deviates by less than 15 degrees
(low deviation range); yellow that the student deviates between 15 and 30 degrees (medium deviation range); and red encodes deviation of greater than 30 degrees (high deviation range). As with the color of the touch map elements, the color of an arrow is calculated by linearly interpolating between the two neighboring ranges of the learner’s deviation value.

Tracking the learner’s palpation position by an infrared marker on the middle finger allows the pattern-of-search map to indicate the deviation of a move before the learner actually palpates. This extra arrow is visually distinct as it is not outlined (see Figure 7-11). The learner is guided through correct pattern-of-search by the expert pattern. The color of the current arrow between the last palpation position and the current hand position also provides guidance by indicating whether palpating at the current hand position would follow or deviate from the expert’s pattern. Reinforcement and correction are provided by arrows of prior palpations being colored green or red.

The touch map and pattern-of-search map are presented in combination (Figure 7-14) to provide the learner with feedback of cognitive and psychomotor components of CBE: palpation completeness, palpation pressure, and pattern-of-search.

7.6 Post-Experiential Feedback

Learning is maximized when feedback and reflection occurs during and after an experience [12]. After a MRIPS-CBE interaction, learners receive feedback which provides a summary of their performance on cognitive, psychomotor, and affective aspects of the CBE.

Feedback is divided into performance in the medical history (cognitive) and affective portions of the exam (Figure 7-15) and performance in the visual inspection and palpation portions of the exam (cognitive and psychomotor; Figure 7-16).
Feedback is quantitative, to motivate learners to improve their scores by improving their exam performance. These quantitative ratings of performance can also be used by educators to more precisely grade portions of the CBE which can not be graded quantitatively without MRIPS: correct pressure, coverage, and following of the pattern-of-search.

7.7 Face Validity of Touch Map and Pattern-of-Search Map Feedback

The face validity of the touch map and pattern-of-search map was established through feedback from novices and experts in an informal evaluation. Face validity indicates that the touch map and pattern-of-search map appears to assist the learner in performing a more complete and correct exam, with respect to palpation coverage, palpation pressure, and pattern-of-search. Whether this feedback does improve learners’ exam skills will be evaluated in Chapter 9. To establish face validity, expert clinicians and 2nd-year medical students provided informal feedback concerning the touch map and pattern-of-search map feedback.

After receiving lecture-based CBE teaching, six 2nd-year medical students performed their first CBE using a version of MRIPS-CBE integrating the touch map visualization (but not the pattern-of-search map). All students reported that the touch map assisted them in palpating the entire breast and in using correct pressure, and felt that receiving this feedback was valuable in the learning process. A portion of students’ responses are shown below, with italicized phrases indicating how each quote relates to coverage or palpation pressure e.g.:

- “Being able to see which areas I had covered, helped me to examine the entire breast.”
- “The map helped me realize how deep I was pressing, and how large an area I covered.”
• “It was good being able to watch and see when I hadn't gone far enough or when my next row was too far over.” [this student is referring to the feedback allowing her to see when she has left too much space between rows of her vertical strip pattern-of-search]

• “I thought the touch map was an excellent way to learn how to do a breast exam. It was nice to feel how deep you are supposed to palpate.”

• “I thought the most important feature was the pressure sensing capability. [Before using MRIPS-CBE] we don't really know how much force to apply and I think this would be very useful if implemented into our curriculum prior to our 3rd yr rotations [in which students are graded on CBE].”

Students’ comments indicate that their psychomotor skill of palpatting with correct pressure was undeveloped before using MRIPS-CBE, and that the touch map feedback assisted in developing this psychomotor skill by allowing them to link the amount of force applied with the level of pressure visualized. Students also indicated that the touch map assisted them in the cognitive task of palpating the entire breast, helping them keep track of the area they had palpated and making sure adjacent palpations were adequately close together.

All students also reported that the multicolor scale of the touch map was easily interpreted, though one student suggested alternately using the progression of colors of the visible light spectrum.

Additionally, four expert clinicians who have each conducted thousands of CBEs informally evaluated the touch map and pattern-of-search map feedback, providing feedback concerning how the real-time visual feedback could benefit students learning the CBE. Portions of the expert’s views are reproduced here, with italics indicating the connection between the expert quote and the coverage, pressure, and pattern-of-search feedback:

• Expert 1: “One of the biggest complaints students have when they finish medical school is that they don’t have faculty or qualified people observe them conducting a
breast exam to the extent that they would like. …This is information that they wouldn’t get otherwise. …This technology allows [students] to get real-time feedback on the quality of their exam and the ability to incorporate that feedback into their learning and improve their examination skills, so I think it could have a significant impact on their ability to perform breast exams in the future.”

- Expert 2: “This seems good for making sure students examine all portions of the breast. I have not seen much focus on pressure before. …I am sure it will help students with their own comfort level in the CBE which can be a hurdle for some students.”

- Expert 3: “Feedback is probably much better than what students normally receive. Particularly with the pressure feedback.”

- Expert 4: “Too often the breasts are provided only a cursory exam. The idea of placement, appropriate pressure and pattern of examination is quite useful as an educational tool. …[The expert expects that] having had experience on the mannequin will give them more confidence and understanding and will enable them to complete an exam on a living patient more efficiently.”

The expert clinicians identified that the touch map and pattern-of-search map visual feedback provided students with feedback on palpation pressure, pattern-of-search, and coverage of the breast which students are not provided in purely physical learning environments. The experts also remarked on the potential of the real-time visual feedback to overcome barriers to learning CBEs, such as student comfort (anxiety) and confidence.
Figure 7-1. A) The breast history portion of the procedural checklist is displayed above the virtual human’s head. In B) and C), items and topics are highlighted as the user asks questions corresponding to the items on the list.
Figure 7-2. The visual inspection portion of the procedural checklist expands to show the three poses required for visual inspection: relaxed with arms at sides, chest flexed with hands on hips, and arms held above head.

Figure 7-3. The procedural checklist also incorporates feedback to aid in the cognitive task of recalling which peripheral areas of lymph nodes should be examined. This also provides high-level feedback as to the completeness of palpation, as areas are highlighted as they are palpated by the user. This highlighting is illustrated in the progression from A) to B) to C). These areas are axillary, infraclavicular, and supraclavicular. The cone of the breast is also highlighted as it is palpated, though more detailed feedback of completeness of palpation of the cone is given by the touch map feedback of Section 7.4.
Figure 7-4. A) The virtual human prompts for empathy. B) If the learner then expresses empathy, e.g. “I understand how hard this must be for you,” the virtual human’s thoughts indicate positive affect towards the learner. C) If the learner ignores the prompt, the virtual human responds to indicate the learner needs to pay more attention to emotional content of the interaction. D) If the learner responds inappropriately, the virtual human’s thoughts express negative feelings towards the learner.
Figure 7-5. When the physical exam begins, A) the virtual human expresses fear. B) If the learner ignores the opportunity to comfort the patient, her response indicates negative affect towards the learner. C) If the learner responds with a patronizing response, e.g. “there is nothing to be scared about,” the virtual human responds with similar negative affect towards the learner.
Figure 7-6. The touch map provides feedback of coverage and use of correct palpation pressure through color-coded visual elements presented in-situ with the virtual human and the haptic interface.

Figure 7-7. The pairing of the color and infrared seeing cameras (hanging above the mannequin) and the haptic interface to the virtual human.
Figure 7-8. A) The boundary of the area required for complete coverage of the breast cone. B) Complete coverage is indicated when this area is filled.

Figure 7-9. Informal correctness of the model is demonstrated by showing that the output of the model fits the expected progression of pressure levels. Shown here is data for a single sensor. This behavior is consistent across all sensors and repeatable across multiple calibration data sets.
Figure 7-10. The color of the feedback provides guidance, reinforcement, and correction of the learner’s palpation pressure through real-time changes in color. A) Light pressure is indicated by a blue color; B) Medium pressure by yellow; C) Deep pressure by green; and D) Too-high pressure by red.

Figure 7-11. A, B) A learner follows an expert’s vertical strip pattern-of-search. C) The pattern-of-search map indicates the learner’s failure to follow a systematic pattern.
Figure 7-12. Modeling pattern-of-search: A) Expert palpation position data contains clusters. B) The centroids of these clusters. C) The resulting expert path after filtering and spline interpolation. E) The vertical strip pattern is then rendered with direction arrows added. F) The same process applied to a spiral expert pattern-of-search. D) The scalar field used to calculate learner deviation from the pattern of (e) is shown.

Figure 7-13. A) The touch map and B) pattern-of-search map and C) the combination of the two for the same exam. D, E) The progression of the combined visualizations.
Figure 7-15. Feedback is provided concerning cognitive elements such as the procedure of visual inspection and cognitive-psychomotor elements such as completeness of palpation, palpation with correct pressure, and following of the vertical strip pattern of search.
Figure 7-16. Summary feedback of affective (“Your use of empathy”) and cognitive (“Your information gathering”) performance. If the learner did not ask about a topic, it appears as red in this list. This list includes items in addition to those in the procedural checklist real-time feedback (Section 7.2). Items on this list, other than the patient concerns, are taken from validated checklists used to grade medical students at the University of Florida’s College of Medicine and at the Medical College of Georgia.
CHAPTER 8
FEEDBACK IN MRIPS-NEURO

8.1 Introduction

This chapter describes the development and evaluation of two visualizations providing feedback to guide cognitive, psychomotor, and affective performance in MRIPS-NEURO.

The thesis states that an interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance will improve users’ psychomotor, cognitive, and affective skills. To prove this statement, we must demonstrate that real-time feedback improves performance in these skill sets. This chapter describes a formal study designed to evaluate whether providing H-Map and Patient Vision feedback improves learners’ performance in these three skill sets.

We have developed two visualizations to provide guiding and motivating feedback in cognitive, psychomotor, and affective components of the cranial nerve exam. The “H-Map” displays a path for the learner to follow during the eye movement test, guiding the psychomotor task of moving the finger or ophthalmoscope in an “H” pattern to test the extremes of the patient’s vision. The “patient vision” visualization allows the learner to view the virtual world through the eyes of the patient. We expect that experiencing the patient’s double vision firsthand will assist learners in the cognitive task of diagnosing the affected cranial nerve (based on eye movements) and improve learners’ perspective taking and concern for the patient, improving their affective performance in the exam.

8.2 H-Map

The H-Map (Figure 8-1) visualizes the uppercase “H” pattern used to assess the patient’s range of eye movements. In this assessment, the learner sweeps his finger or
the ophthalmoscope light in the “H” shape, testing the lateral extremes of the patient’s vision and at each lateral extreme (left and right) testing the vertical extremes of the patient’s vision. If the patient can not move his eyes (or one of his eyes) through the entire “H” pattern, this indicates a cranial nerve 3, 4, or 6 disorder. For example, if the patient has a CN3 affected left eye, he can not adduct the left eye towards his nose. The eye movements are the most important test for diagnosing cranial nerve palsies that restrict eye movements, making this test important for learners to perform correctly. The H-Map guides learners to ensure they are testing the extremes of the vision.

This H-Map is implemented as colored targets (colored quadrilateral primitives) connected by two vertical and one horizontal bar (also colored quads). The targets represent the positions the ophthalmoscope or hand should reach in order to elicit the six cardinal extremes of the patient’s eye movements. The position of the targets varies with the depth of the ophthalmoscope or hand tool from the virtual human’s eyes. Targets are placed at the intersection of rays cast from the unaffected right eye of the virtual human with the XY plane placed at the depth of the tool. Each of the targets corresponds to casting of a ray oriented in one of the six extreme rotations of the eye.

Intersecting the targets with the gesture (hand) tool or ophthalmoscope tool changes the color to green to indicate the eye movement extreme corresponding to that target has been tested.

Feedback on this design was received through informal testing with a neurologist and four neurology residents. These experienced performers of the neurological exam indicated that they performed the eye movements test close in depth to the patient’s head, approximately 1 ft. or less. This allowed them economy of movement, and they
could focus on the patient’s eyes while keeping their finger/ophthalmoscope in their primary vision. Based on this feedback, we altered the H-Map to indicate to the learner a range of depth that would provide this economy of movement. The horizontal and vertical bars of the H-Map change to a green color when the ophthalmoscope or hand is approximately 5-12 inches in depth from the patient’s eyes (Figure 8-7).

8.3 Patient Vision

8.3.1 Feedback Goals

The patient vision feedback allows learners to view the virtual world through the eyes of the patient, providing learners with firsthand experience of life with double vision. The goal of this experience is to motivate learners to engage in social perspective taking, imagining what it is like to be the patient. When a person engages in perspective taking, he first considers what he knows of the other person’s knowledge, senses, and experiences. This leads him to an affective understanding of the other – understanding the other’s emotions and state-of-mind [139]. The outcome of this process is expression of this affective understanding, through empathy and concern for the other.

Thus the goal of the patient vision feedback is to aid learners in understanding how the patient’s double vision affects the patient’s life and emotional state. We expect the outcome of this understanding to be increased consideration of the patient’s safety, e.g. is it safe for the patient to drive home from the clinic, and expression of empathy – improving the learner’s affective performance in MRIPS-NEURO.

We are motivated to incorporate the patient vision feedback by prior work which allowed a learner to relive a clinical breast exam from within the body of the patient, and
demonstrated impact on learner’s self-awareness of his affective performance in the CBE.

8.3.2 Prior Work in Motivating Perspective Taking

Raij developed an after-action-review experience, virtual social perspective taking (VSP), which sought to improve medical students’ use of perspective taking in the clinical breast exam scenario [85]. Learners performed CBE of a virtual human patient using MRIPS-CBE, and then relived their interview and exam from within the body of the virtual human. The learner looked through the virtual human’s eyes, seeing what the virtual human saw during the exam – the virtual world and the learner. As the virtual human, the learner heard what the learner had said during the exam and was asked to speak what the virtual human spoke during the exam. The learner embodied the avatar of the virtual human and controlled the pose of the avatar’s head. To emphasize that the learner was reliving the experience in the body of the virtual human, the learner could see the movements of his avatar in a virtual mirror (Figure 8-2).

Raij and I conducted a study of 16 medical students, residents, and clinicians at the Medical College of Georgia to evaluate the impact of VSP on perspective taking and empathy. Participants rated their affective performance along the dimensions of perspective taking and empathy before and after the VSP. Ratings decreased after VSP, demonstrating that reliving the experience from the patient’s point of view helped learners become more aware of their affective performance. The VSP feedback motivated reflection, which leads to change. Participants indicated that they would change how they approached perspective taking and empathy in future patient interactions, but change was not specifically measured in this study.
8.3.3 Patient Vision Feedback

Patient vision is a simulation of what the virtual human with cranial nerve disorder sees. By wearing an HMD, the user is able to see through the patient’s eyes and experience the double vision and incomplete range of eye movement experienced by the patient. This visualization is targeted to improve the learner’s cognitive performance in diagnosing the cranial nerve disorder and to improve the learner’s affective performance by motivating perspective taking which is expected to increase empathy and concern for the patient’s safety.

The patient vision feedback is a novel simulation of the experience of double vision and a cranial nerve disorder. Prior work has developed visual simulation of other vision disorders: myopia and effects of laser surgery on myopic vision [140], recreating an individual’s vision by scanning the retina [141], and multitexturing to simulate glaucoma and diabetic retinopathy [142].

The learner literally sees through the eyes of the patient, as the virtual world is rendered from two cameras having the position and orientation of the virtual human’s eyes (Figure 8-3). The virtual human’s left eye image is presented to the user’s left eye and the virtual human’s right eye image is presented to the user’s right eye. Just as in a patient with double vision, the job of fusing the two images is left to the user’s brain. If the virtual human is seeing double, then the user will be unable to fuse the two images and as a result will also see double. During the neurological exam, the virtual human’s behaviors allow the user to experience what a patient with double vision sees during a neurological exam. When the patient is asked to follow the doctor’s finger with his eyes, the user is able to determine in what fields of vision (e.g. left, right, far, close) the patient sees double. The patient vision also reflects other changes in the patient’s vision during
the exam. For example, when the patient is asked to close or cover one eye (e.g. in the visual acuity and finger counting tests), the covered eye’s image is not rendered, providing the user with vision in only one eye.

Initially we had experimented with combining the left and right eye images into a compositied image, by alpha blending the left eye on top of the right eye with alpha = 0.5 (Figure 8-4). This approach provided a method for experiencing the patient’s double vision displayed directly on the non-stereoscopic large screen display. During the exam, the virtual human would say “let me show you what I see” and the learner’s view of the virtual human would be replaced by the virtual human’s view of the virtual world. While this approach could benefit the affective component of perspective taking, it is unable to assist the cognitive task of diagnosing which cranial nerve is affected, because the left and right eyes can not be distinguished.

To afford diagnosis during patient vision, a stereo display, such as an HMD, is required. For this reason, and to facilitate the study design of Section 8.4, the implementation was altered to render each virtual human eye view to its respective HMD screen (left, right).

8.4 Evaluating the Impact of Feedback on Cognitive, Psychomotor, and Affective Performance

We conducted a user study to evaluate the impact of the H-map and patient vision feedback on learners’ cognitive, psychomotor, and affective performance.

8.4.1 Study Design and Procedure

To directly evaluate the impact of the patient vision feedback on participant performance in the cognitive task of diagnosing the CN disorder and in the affective task of perspective taking, participants were divided into two groups. Group A experienced
the patient vision feedback before examining the patient and Group B did not experience patient vision before examining the patient.

The procedure is shown in Figure 8-5. Two participants arrived at a time. Each participant completed a background survey assessing their experience with the neurological exam.

The experimenter then explained to both participants how to talk to the virtual human and how to use the Wiimote to manipulate the virtual ophthalmoscope, hand, and eye chart tools. The experimenter pointed out the series of icons on the left hand side of the screen which illustrated the tests that could be performed on the virtual human (Figure 8-7). Participants were told the patient vision feedback would let them experience the patient’s double vision during the exam and that they should try to figure out which eye and cranial nerve is abnormal during the patient vision. Participants were told that the H-Map feedback would appear during the eye movements test. They were told that the size of the H would decrease as they moved the hand/ophthalmoscope closer to the patient, that the H would change color when the hand/ophthalmoscope was an “optimal” distance from the patient, and that the targets represented the six cardinal extremes of the patient’s eye movements. Participants were not told that they were required to follow the H-Map, though all participants did follow the H-Map.

After receiving instruction, Participant A donned a stereoscopic HMD displaying the virtual exam room and Participant B wore a hat augmented with infrared markers used to track her head pose. Participant B viewed the virtual human and exam room rendered on the non-stereoscopic large screen display. By tracking her head pose, the
virtual scene is rendered from her perspective, as if the display was a window into the virtual world (Figure 8-6).

Participant B performed a medical history and exam of a virtual human with a CN6 affected left eye. In this exam Participant B received the “H-map” feedback during the eye movement test. During this exam, participant A viewed the virtual world, including the tools manipulated by Participant B, through the eyes of the virtual human. This allowed Participant A to experience the patient’s double vision during an exam, allowing Participant A to evaluate the severity of the patient’s double vision and the patient’s eye movements. Participant A was a passive observer; the movement of virtual tools was controlled by Participant B and the speech, actions, and eye movements that Participant A experienced were controlled by the virtual human simulation.

After Participant B completed the history and exam of the virtual human patient with CN6 disorder, Participant A removed the HMD and completed the post-patient-vision survey (Appendix H).

Participant B then performed an exam (without taking a medical history) of a virtual human with a CN3 affected left eye. In this exam, the H-Map feedback was not provided. Participant B then completed a post-exam survey (Appendix I).

The participants then switched roles and repeated the procedure, with Participant B experiencing the patient’s vision and Participant A performing a history and exam of the CN6 virtual human patient followed by an exam of the CN3 virtual human patient.

The counterbalanced design of Group A and Group B’s procedures allows us to evaluate the impact, on cognitive and affective performance, of experiencing a patient’s abnormal vision before examining the same patient. By having each participant
examine both a CN6 and a CN3 patient, this design additionally allows us to make within-subjects comparisons of the completeness and efficiency of the eye movement tests when H-Map feedback is provided (CN6 patient) and when H-Map feedback is not provided (CN3 patient). The eye movement test required for the CN6 and CN3 patients is equivalent.

8.4.2 Population

Eighteen 2nd-year medical students at the University of Florida’s College of Medicine participated. All participants had experience in neurological exams of standardized or real patients, with four participants having 1-5 experiences, eight having 6-10 experiences, and six having more than 10 experiences. It is not known whether participants had previous exposure to CN3 or CN6 in human patients. Nine participants were enrolled in each of Group A (patient vision before exam) and Group B (patient vision after exam).

8.4.3 Metrics

8.4.3.1 Evaluating cognitive and affective performance

After completing the patient vision experience, participants were asked to diagnose the cranial nerve and eye affected based on what they saw through the patient’s eyes. Participants are also asked to describe how the participant felt the double vision affected the patient’s everyday life. This survey assesses the impact of patient vision on performance in the cognitive task of diagnosing the cranial nerve disorder and the affective task of perspective taking.

After examining the CN3 and CN6 patients, participants were asked to provide a diagnosis of the cranial nerve and eye affected in each of two patients. This was used to evaluate cognitive performance. Participants were also asked to list any concerns
they would like to relay to the patient or the patient’s family. This item was also used to
evaluate perspective taking.

In evaluating the two free response items in the post-patient-vision and post-exam
surveys, we looked for expression of participant concern for patient safety. These
consisted of expressions that the patient should not engage in specific tasks that have
become dangerous to perform due to the patient’s double vision. Specifically, we
expected to find instructions to the patient that he should not drive a vehicle. The
patient stated early in the exam that the double vision had started when he was driving
home from work. If asked, the patient also states during the exam that he drove to the
doctor’s office and would be driving home. We specifically focused on driving as an
aspect of patient safety because driving is an everyday task that all participants knew
the patient performed. Based on the patient’s double vision, the participant should do
her best to dissuade the patient from driving due to the danger to the patient and others
[52]. Expression for patient concern was also evaluated by observing video of the
participants’ history taking and exam, i.e. to determine if participants verbally instructed
the patient that he should not drive or expressed concern about how the patient would
get home from the doctor’s office.

8.4.3.2 Evaluating psychomotor performance

The psychomotor component of the exam is the eye movement test. The impact
of the H-Map feedback on the completeness and efficiency of the eye movement test
was measured by recording the position and orientation of the virtual tools at a minimum
of 30 Hz.

Completeness of the eye movement test was evaluated as the difference of 1) the
angles of the six cardinal eye movements elicited by the participant and 2) the
maximum angles of the six cardinal eye movements. The yaw and pitch elicited by the participant was treated as a two-dimensional vector. The maximum yaw and pitch was also treated as a two-dimensional vector. The difference between the participant-elicited yaw and pitch and the maximum yaw and pitch was calculated as the Euclidean distance between these two vectors. For example, in CN3, the right-down extreme is \((\text{yaw} = 10, \text{pitch} = 10)\). If the participant elicited \((\text{yaw} = 10, \text{pitch} = 7)\), the distance for this extreme is calculated to be 3 degrees. The differences at each of the six extremes were summed for each participant. A sum of zero indicated the most complete eye movement test possible.

As described in Section 8.2, medical professionals experienced at neurological examination perform the eye movement test with the finger or ophthalmoscope held one foot (or less) in depth from the patient’s eyes. This affords an economy of motion of the practitioner’s hand/arm and allows the practitioner to view the patient’s eyes up close while keeping the practitioner’s finger (or the ophthalmoscope) within the practitioner’s primary vision. The horizontal and vertical bars of the H-Map visualization became colored green when the virtual hand or ophthalmoscope was held at a level of depth of \(\sim 5''\) to \(\sim 11''\). These depths are indicated to be approximate because the implementation calculated depth in centimeters. The depths in centimeters were indicated by an expert neurologist testing MRIPS-NEURO. An efficient eye movement test would test all six extremes at such a depth. To evaluate efficiency of each participant’s eye movement test, we counted the number of extremes that were tested within the efficient depth range. We also calculated the average depth at which the six extremes were tested and the standard deviation of the depths at which the six extremes were tested. This
was performed for each of the CN6 exam in which the H-Map was provided and for the CN3 exam in which the H-Map was not provided.

8.4.3 Hypotheses

Hypotheses concern the impact of the Patient-Vision feedback on improving the affective task of perspective taking and the cognitive task of diagnosis, as well as the impact of the H-Map feedback on the completeness and efficiency of the psychomotor task of testing patient eye movements.

- **Hypothesis Patient-Vision Improves Affective**: Participants experiencing the patient vision feedback (CN6) before performing an exam of the patient (CN6) will exhibit increased concern for patient safety, expressed verbally to the patient or written in the post-patient-vision or post-exam surveys.
  - Null hypothesis: Participants in Group A and Group B will exhibit no difference in verbal or written concern for patient safety.

- **Hypothesis Patient-Vision Improves Cognitive**: Participants experiencing the patient vision feedback (CN6) before performing an exam of the CN6 patient will more often correctly diagnose the cranial nerve disorder of the CN6 patient.
  - Null hypothesis: There will be no difference in the number of participants in Group A and Group B who diagnose the CN6 patient correctly.

- **Hypothesis H-Map Improves Psychomotor Completeness**: Participants will perform more complete eye movement tests when the H-Map is provided than when the H-Map is not provided.
  - Null hypothesis: Participants’ eye movement tests of the CN6 and CN3 patients will be equivalently complete.

- **Hypothesis H-Map Improves Psychomotor Efficiency**: Participants will perform more efficient eye movement tests when the H-Map is provided than when the H-Map is not provided.
  - Null hypothesis: Participants’ eye movement tests of the CN6 and CN3 patients will be equivalently efficient.

We do not expect to improve participants’ psychomotor skills from a single interaction, thus we evaluate the H-Map on its ability to elicit more complete and
efficient performance when the H-Map is present. Prior work has demonstrated that similar psychomotor tasks can be learned through similar mixed reality visualizations of tool or hand trajectories [143]. If learners demonstrate more accurate (complete) and efficient eye movement tests when the H-Map feedback is provided than when the H-Map feedback is not provided, we can infer that repeated practice of the eye movement test with the H-Map feedback will lead to more accurate and efficient eye movement tests.

8.4.4 Results and Discussion

Two participants, one in each group, expressed that they did not experience double vision at any time during the patient vision feedback. We were unable to determine whether this was due to malfunctioning of the HMD or a peculiarity of the participants’ vision. It is possible for the HMD to automatically toggle between non-stereoscopic (left eye duplicated for both left and right eyes) and stereoscopic display; however, the experimenter tested the HMD before each participant and did not note any problems in this regard. However, because these participants did not actually experience the virtual patient’s double vision, and were evenly split between the two groups, they were removed from analysis for Hypothesis Patient-Vision Improves Affective and Hypothesis Patient-Vision Improves Cognitive.

8.4.4.1 Hypothesis Patient-Vision improves affective. Experiencing patient vision increases concern for patient safety: accepted

Affective performance was evaluated as expressed (verbal or written) concern for patient safety, i.e. informing the patient that he should not drive a vehicle. Significantly more participants who experienced patient vision before examining the virtual human patient (Group A) expressed concern that the patient should not drive than did
participants who did not experience patient vision before the exam (Group B). Seven of eight participants in Group A expressed this concern vs. zero of eight participants in Group B. This is significant at $p < 0.005$ by Fisher’s exact probability test.

Almost all participants who experienced patient vision before the exam expressed concern that the patient should not drive, with five participants expressing this on the post-patient-vision survey, one participant (of the five) directly telling the patient during the exam, and three participants expressing this on the post-exam survey (one participant in this group had previously indicated this in the post-patient-vision survey). None of the participants in Group B, who performed the exam before experiencing patient vision, expressed concern that the patient should not be driving. All patients were equally primed to think about driving, as the patient first began to experience double vision while he was driving home from work: “I was driving home from work and all of the sudden the lines on the road started to cross.” However, only those participants who had seen through the eyes of the patient before assuming the role of the doctor expressed concern for the patient’s safety. This indicates that literally providing participants with the patient’s visual perspective caused the participants to engage in perspective taking. The participants experiencing Patient-Vision considered the patient’s feelings and the impact of the double vision on the patient’s life, and developed a concern for the patient’s safety that later (as the doctor) allowed them to identify driving as a danger to the patient. Thus participants’ affective performance was improved by providing them with the patient vision feedback prior to their examination of the virtual human patient.
8.4.4.2 Hypothesis Patient-Vision improves cognitive. Experiencing patient vision aids diagnosis of CN disorder: rejected but with a positive result

Nearly all participants in both groups were able to correctly diagnose both the CN3 and CN6 virtual patients through physical examination (9 of 9 in Group A and 8 of 9 in Group B). This leads us to reject Hypothesis Patient-Vision Improves Cognitive, as performance in the cognitive task of diagnosis was equivalent for those participants who received patient vision prior to the exam and those participants who did not receive patient vision prior to the exam. As participants in both groups performed well in this task, we can not conclude whether the patient vision improved, or did not improve, performance.

However, there is evidence that patient vision alone is adequate for diagnosing the cranial nerve disorder. Eighty-one percent of participants correctly diagnosed the cranial nerve disorder while viewing Patient-Vision. Seven of the eight participants in Group A were able to correctly diagnose the virtual patient with CN6 palsy from experiencing patient vision alone (i.e. before conducting an exam of the patient). Additionally, 6 of 8 participants in Group B were able to diagnose CN6 from the patient vision feedback. Although Group B had previously examined a virtual human patient with CN6, they had most recently examined a virtual human with CN3 and were not told what disorder they would experience with patient vision (it could have been CN3, CN6, or other). Overall, 13/16 participants were able to correctly diagnose CN6 from experiencing the patient’s vision. This is a significant percentage of the participants (one-way chi-square test: $X^2 = 5.1, p < 0.05$). This result shows that viewing a (virtual) neurological exam through the eyes of a virtual human patient with cranial nerve palsy
provides the information needed to diagnose the cranial nerve disorder affecting the virtual human’s eye movements.

This is a positive result for the impact of Patient-Vision on cognitive performance: experiencing the patient’s double vision aided in the diagnosis of the cranial nerve disorder. The rejecting of Hypothesis Patient-Vision Improves Cognitive is caused by a too-restrictive wording of the hypothesis and a ceiling effect in participants’ diagnostic performance. While Patient-Vision is sufficient for diagnosing the cranial nerve disorder, we were unable to measure the impact of Patient-Vision on the correctness of diagnosis reported after an exam of the patient. To determine if patient vision is a significant factor in arriving at a correct diagnosis after additionally performing an exam of the patient, we should revisit this experiment with a more difficult-to-diagnose cranial nerve disorder.

In addition to evaluating the cognitive impact of Patient-Vision, the finding that 17 of 18 participants arrived at correct diagnoses provides additional evidence of the content validity discussed in Section 5.5. In the prior study of Section 5.5, 12 of 14 participants were able to use MRIPS-NEURO to diagnose one cranial nerve disorder. In this study, 17 of 18 participants were able to use MRIPS-NEURO to correctly diagnose two cranial nerve disorders. This result strengthens our claim of the content validity of MRIPS-NEURO: MRIPS-NEURO simulates a neurological exam with abnormal findings to fidelity sufficient to allow learners to arrive at a correct diagnosis.

8.4.4.3 Hypothesis H-Map improves psychomotor completeness. H-Map visualization results in a more complete eye movements test: rejected

We expected participants to perform a more complete eye movements test, eliciting all extremes of the patient’s eye movement, when the H-Map was provided.
This expectation was not supported by the data. Five participants performed a more complete eye movements test with the H-Map present (avg. of 4.8 ± 2.8 degrees more complete with H-Map). Four participants had equivalent tests with and without the H-Map present (total difference < 1.0 degree). The remaining nine participants performed better without the H-Map present (avg. of 4.3 ± 1.3 degrees less complete with H-Map). A t-test comparing the magnitude of the difference for the 5 who performed better with H-Map and the 9 who performed worse with H-Map revealed no significant difference: \( t(12) = .61, \text{ ns} \). A within-subjects paired samples t-test revealed no significant improvement from viewing the H-Map: \( t(17) = 0.86, \text{ ns} \). We conclude that the H-Map did not assist participants in eliciting the extreme eye movements of the patient.

In investigating why the H-Map did not improve completeness of participants’ exams, we looked for specific eye movement extremes that caused difficulty for participants and reviewed participant comments. The average differences listed in the previous paragraph seem small considering they are sums over the six extreme eye rotations. However, more in-depth analysis reveals that these differences are primarily due to missing two of the six extremes: the right-up extreme and the right-down extreme.

With the H-Map, examining the CN6 patient, 12 participants missed the right-up extreme by an average of 2.9 degrees and 12 participants missed the right-down extreme by an average of 3.0 degrees. Without the H-Map, examining the CN3 patient, 12 participants missed the right-up extreme by an average of 4.0 degrees. It was not the same 12 participants missing all three extremes, though there were some participants who missed two.
Participants’ difficulty eliciting the right-down and right-up extremes does not appear to be linked to the cranial nerve disorders examined. A CN6 affected eye has normal movement in the right-down and right-up extremes, while a CN3 affected eye has abnormal movement in these extremes.

Notably, all participants held the Wiimote in their right hands. As the right-up and right-down extremes required moving the Wiimote to the left of the participant's body, the handedness of the participant may have played a role. By examining participant feedback and behavior, we have identified two other potential factors linked to participant skill and system design.

During the exam, participants reported difficulty in judging the depth of the ophthalmoscope or hand from the virtual human’s head. This was likely due to the virtual world being displayed on a non-stereoscopic display. The lack of stereo depth cue appeared to cause participants difficulty in following the H-Map. While following the H pattern, participants would tilt the top of the Wiimote towards the screen, causing the depth of the virtual hand/ophthalmoscope to become closer to the patient. This in turn caused the H-Map to become smaller. Participants appeared to perceive the decrease in size of the H-Map to indicate the H-Map had moved farther away from them, and in turn continued moving the hand/ophthalmoscope closer to the patient, “chasing” the H with the virtual hand/ophthalmoscope. This occurred because of the lack of stereo depth cues and two other factors: (1) Participants seemingly forgot prior instruction that the H appeared at the same depth as the virtual hand/ophthalmoscope. (2) Participants did not use appropriate movements of the Wiimote – e.g. when having a human patient follow the light on the ophthalmoscope, tilting the ophthalmoscope’s tip
towards the patient would cause the light to aim at the floor and the patient would no longer see the light they were instructed to follow. We believe that providing a stereoscopic display would eliminate factor (1).

Eliminating factor (2) appears to be more complex. The reason for these inappropriate movements of the Wiimote appears to be that the participant tended to move the Wiimote in orbit about his or her own body – instead of more appropriately moving the Wiimote as though it was in orbit around the virtual human patient’s head. We believe this is partially due to participant inexperience i.e. lack of skill in this psychomotor task, i.e., novices might make these same egocentric movements with a real ophthalmoscope. However, participants’ behaviors raises questions for future work: do participants have difficulty in mapping the movement of the Wiimote to the movement of the virtual tool perceived to be beyond the display surface? Perhaps the approach of adding a physical mannequin head would decrease the cognitive load imposed by this mapping.

While the H-Map in its current form does not appear to guide learners to perform more complete eye movement tests, further evaluation reveals that practice with the H-Map has the potential to improve the efficiency of participants’ exams. Potential solutions must mitigate the high variance in the tracked depth position, reducing the impact of user errors such as pointing the wiimote at the screen. One such solution is to track the position of the wiimote as the centroid of the wrist. In the study, the position was tracked as the tip of the wiimote because this corresponded closely with the tip of the virtual hand’s finger – which the patient was to follow with his eyes. However, the offset between the tip of the wiimote and the user’s wrist can be estimated with
reasonable confidence, making it possible to use the tip of the Wiimote as the target for
the patient’s eyes to follow and the position of the wrist as the depth of the H-Map.

8.4.4.4 Hypothesis H-Map improves psychomotor efficiency. H-Map visualization
results in a more efficient eye movements test: accepted

Following the H-Map allowed participants to perform more efficient eye movement
tests than when the H-Map feedback was not provided. An efficient eye movements
test would test all six extremes within the “optimal” depth range described in 8.4.3.2,
and would test all six extremes at the same depth, i.e. with minimal deviation in depth.

Sixteen of the 18 participants examined more of the six extremes within the
“optimal” depth range when the H-Map feedback was provided than when the H-Map
was not provided. This is a significant majority by one-way chi-square ($X^2 = 9.4, p <
0.005$). On average, participants tested significantly more extremes within the optimal
depth range when viewing the H-Map: $4.7 \pm 1.3$ extremes vs. $1.7 \pm 1.8$ extremes without
the H-Map. This difference is significant by paired samples t-test: $t(17) = 6.3, p < 0.005$.

Participants also examined the six extremes of the patient’s vision with
significantly less variance in depth when they were able to follow the H-Map than when
the H-Map was not displayed. The standard deviation of the depths at which the six
extremes were tested: with H-Map: $2.4 \pm 1.7$ in. and without H-Map: $3.6 \pm 1.8$ in.

The H-Map allowed these novice learners of the neurological exam to perform an
eye movements test of similar efficiency as an expert. Without the H-Map present,
participants performed this psychomotor task using a previously learned, less efficient
method. To illustrate this, without the H-Map some participants performed the eye
movements test at a depth that required them to hold the Wiimote out at arms length,
with the virtual hand/ophthalmoscope no longer appearing on the screen. If novice
users of MRIPS-NEURO were to practice the eye movements test repeatedly with the H-Map feedback, we would expect these novices to improve in efficiency of this psychomotor task.

8.4.5 Conclusions

In this study, we evaluated the impact of the patient vision and H-Map feedback on participants’ performance in the affective, cognitive, and psychomotor components of the neurological exam.

The patient vision feedback was targeted to the affective task of taking the patient’s perspective and a resulting expression of concern for the patient’s safety. Participants who experienced the patient’s vision before examining the patient expressed concern that it would be dangerous for the patient to drive. These participants demonstrated increased affective performance in perspective taking over the participants who did not experience the patient vision before examining the patient.

Experiencing the patient’s vision before examining the patient was also expected to aid in the cognitive task of diagnosing the patient’s cranial nerve disorder. We expected participants who experienced patient vision to arrive at a correct diagnosis from the exam more often than those participants who did not experience the patient’s vision prior to the exam. This was not shown to be true as nearly all participants in both groups arrived at a correct diagnosis. However, a significant majority of the participants were able to diagnose the disorder solely from experiencing the patient vision, demonstrating that the patient vision feedback does aid in the cognitive task of diagnosing the cranial nerve disorder.

The patient vision feedback was successful in its goals of improving affective and cognitive performance. In the right scenario, this feedback may be a powerful method
of emphasizing both affective and cognitive aspects of the neurological exam to novice learners. For example, in the scenario of an elderly patient with dementia who is unable to adequately express his vision problem, patient vision feedback could be used to emphasize the importance of certain tests, e.g. fundoscopic examination, as well as provide the learner with an emotional grounding that would aid in making an affective connection with the patient, i.e. achieving rapport with a patient who has difficulty communicating.

The H-Map feedback was expected to improve the completeness and efficiency of the psychomotor portion of the neurological exam: the eye movements test. The presence of the H-Map improved the efficiency of participants’ eye movements tests, but did not impact the completeness of these tests. Possible causes highlighted potential system design improvements such as providing a stereoscopic display and a less cognitively demanding merging of the virtual and physical spaces.

Overall, the introduction of the two feedback elements into MRIPS-NEURO had a positive impact on participants’ affective, cognitive, and psychomotor performance. After demonstrating that real-time feedback elements can positively impact performance in these skill sets, we turn our focus to learning. We next examine whether repeated practice with MRIPS-CBE while receiving real-time and post-experiential feedback leads to affective, cognitive, and psychomotor learning in the clinical breast exam.
Figure 8-1. Progression of the H-Map visualization as the learner performs the eye movement test.
Figure 8-2. In the VSP feedback experience of Raij et al, the learner: A) performs CBE of a virtual human using MRIPS-CBE, and then B) relives the experience in the avatar of the virtual human.

Figure 8-3. Patient-Vision with a left eye affected by CN3 (A, C, E) and CN6 (B, D, F). Looking: A,B) straight ahead; C,D) down-left; and E,F) to the right.
Figure 8-4. The initial implementation of patient vision alpha blended each eye’s image to present double vision on a non-stereoscopic display. However, this approach makes it difficult to distinguish which image corresponds to the left and right eyes.
Figure 8-5. Study procedure. Group A received patient vision feedback before examining the virtual human patients. Group B did not receive this feedback before the exams. Group B received patient feedback after the exams in order to provide both groups with an equivalent set of experiences.
Figure 8-6. Physical setup of study. The standing participant is performing an exam of the virtual human while the sitting participant wears an HMD to view the exam through the virtual human’s eyes.
Figure 8-7. Views during the exam: A) view of the participant performing the exam; B) view of the participant experiencing patient vision. Image B is sized to enable the reader to experience the double vision. If the reader begins to cross his eyes while focusing on the finger, he should see double – two fingers.
CHAPTER 9
LEARNING, TRAINING TRANSFER, AND IMPACT OF REAL-TIME FEEDBACK IN MRIPS-CBE

This chapter describes a user study, Study MRIPS-Learning, which evaluated learning of cognitive, psychomotor, and affective skills within MRIPS-CBE and transfer of learned skills to the real-world interpersonal scenario of performing a CBE of a standardized human patient (SP). Participants completed a baseline evaluation consisting of a CBE of an SP. Participants then completed a series of three MRIPS-CBE interactions including real-time and post-experiential feedback. Improvement from the repetitive practice with MRIPS-CBE was measured by a second, subsequent CBE of an SP. Learning was evaluated as improvement throughout the three MRIPS-CBE interactions. Training transfer was measured as improvement from the baseline (pre-test) SP interaction to the subsequent post-test SP interaction. Learning and training transfer were demonstrated for cognitive, psychomotor, and affective components of the CBE. Study MRIPS-Learning also evaluated the impact of real-time feedback on performance, through comparison to historical control groups. Results of these comparisons demonstrated significant improvement in cognitive, psychomotor, and affective task performance as a result of receiving real-time feedback.

Collaborators: Educational technology professor Rick Ferdig was consulted in designing the study. Medical collaborators Scott Lind and Brenda Rossen recruited participants for the study. Scott Lind, Brenda Rossen, Andy Laserno, James McLoughlin, Jamison Weir, Steven Blackwood, and Amrew El Alamad assisted with evaluating participants’ SP psychomotor performance through video review. Those video raters and Joanna Lind, Carson Kisner, and Jennifer Carrick assisted with video review to evaluate affective performance in SP interactions.
Personal contributions: I was the primary designer of the study and performed all analysis other than the video review of SP interactions.

Relevance to thesis: The thesis states: “Interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance (MRIPS) improves users’ scenario-specific psychomotor, cognitive, and affective skills. Skills improvement transfers to the real-world interpersonal scenario being simulated, demonstrated as improved performance in the real-world interpersonal scenario.” Study MRIPS-Learning directly evaluates the thesis statement for the CBE scenario by evaluating learning, training transfer, and the impact of feedback on performance.

9.1 Introduction

This chapter describes a user study, Study MRIPS-Learning, that investigated learning (retained skills improvement) in MRIPS-CBE and transfer of learned skills to the real-world interpersonal scenario of CBE of a human patient. Study MRIPS-Learning sought to accomplish four tasks, each with a corresponding meta-hypothesis (formal hypotheses concerning individual measures of performance are given in Sections 9.5 through 9.7).

- Determine what learning occurs in users of MRIPS-CBE.
  - Meta-hypothesis: Participants will improve in cognitive, psychomotor, and affective performance throughout repetitive practice with MRIPS-CBE. Performance in these tasks will significantly improve from the first MRIPS interaction to the third MRIPS interaction.

- Determine whether improvement in skills within MRIPS-CBE transfers to the real-world, in the form of improvement in performance in CBE of human patients.
  - Meta-hypothesis: After practice with MRIPS-CBE, participants’ cognitive, psychomotor, and affective performance in CBE of an SP will have
significantly improved in relation to baseline levels taken before practice with MRIPS-CBE.

- Determine whether the presence of real-time feedback causes learners to significantly outperform past users of MRIPS-CBE (without real-time feedback) in cognitive, psychomotor, and affective tasks.
  
  o Meta-hypothesis: Participants in Study MRIPS-Learning will perform significantly better in cognitive, psychomotor, and affective tasks (for which real-time feedback was provided) than prior users of MRIPS-CBE (participants in Study MRIPSx2, Section 4.3) who did not receive real-time feedback.

- If learning occurs, determine whether skill sets are learned concurrently; or, whether one skill set must be maximized before performance in other skill sets can improve.
  
  o Meta-hypothesis: We will not observe the following: performance in a single skill set requires maximization before the other skill sets can improve.

To accomplish these tasks, we conducted a study with novice medical students at the Medical College of Georgia. Participants were evaluated in a CBE of a standardized human patient (SP), then received a “simulation-based curriculum” consisting of three MRIPS-CBE interactions, and were finally re-evaluated in a CBE of an SP.

9.2 Study Design

9.2.1 Evaluating Learning and Training Transfer

The MRIPS-Learning study procedure is visualized in Figure 9-1. The novice medical students recruited for Study MRIPS-Learning had received lecture-based learning (textbooks, power-point, and expert demonstration with silicone model) of CBE, and had no experience with CBE of standardized patients (SPs) or real patients. The study procedure was as follows:

- Participants were consented and completed a background survey assessing their experience with CBE.
Immediately following the consent and survey, participants performed a CBE of an SP, the SP-Pretest. The SP wore a silicone “breast vest,” incorporating a simulated mass, to evaluate participants on their ability to find breast masses. Video of this CBE was later reviewed by medical educators and myself to establish a baseline for the student’s performance in the cognitive, psychomotor, and affective components of CBE. CBE of an SP was chosen as the real-world interpersonal scenario in which to evaluate training transfer, because:

- SPs have previously been validated for evaluating clinical skills, providing an experience equivalent to a real patient [16].
- SPs provide each participant with an equivalent experience. The SPs were trained to answer participant questions with responses taken from a script (the database of responses used by MRIPS-CBE). While a real patient might describe an ailment slightly differently to each participant, the SP provided the same response to all participants. Providing participants with nearly identical experiences in the SP Pre-test and SP Post-test helps to maintain internal validity of the study – e.g. a participant does not have a “harder” SP in the SP Pre-test than in the SP Post-test (or vice versa), which would skew the delta measured between baseline and final evaluations.
- Real patients with a breast mass were not available for the study. Evaluating participants’ ability to detect breast masses would have required the real patient to wear a “breast vest,” as the SP did; this would reduce the main benefit associated with the real patient, namely the palpating of the patient’s breasts.
- Real patient interactions could not be videotaped (for later review by experts and the experimenters) because of privacy concerns related to the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule. Video recording participants’ CBEs allowed multiple experts to rate the participants’ performances without the logistical restriction of having multiple experts present at the time of the CBE.

Approximately one week after the SP Pre-test, participants completed the first of three MRIPS-CBE interactions. Two MRIPS-CBE interactions followed with approximately one week between interactions. The schedule of the MRIPS interactions had to be approximate in order to accommodate medical student schedules. One week between interactions is a standard time used in repetitive learning studies [63]. Three practice opportunities each spaced approximately one week apart were chosen because this created a MRIPS “curriculum” of similar length to existing clerkships (2-3 weeks in length) used to teach intimate exams in medical curricula. The content of the three MRIPS-CBE interactions was identical, i.e. the appearance, symptoms, and concerns of the virtual human patient were the same. Each MRIPS-CBE interaction incorporated the real-time and post-experiential
feedback of cognitive, psychomotor, and affective performance described in Chapter 7.

- Approximately one week after the third MRIPS-CBE interaction, participants completed a CBE of an SP, the SP Post-test. The content of this SP interaction was the same as the content of the SP Pre-test interaction. This enabled us to evaluate improvement due to repetitive practice with MRIPS-CBE as the delta in performance between SP Pre-test and SP Post-test. Participants received no education in CBE between SP Pre-test and SP Post-test other than the MRIPS-CBE interactions. The time of one week between the end of treatment (the third MRIPS-CBE interaction) and post-test is standard in evaluating whether treatment to post-test improvement is due to skills improvement and retention (i.e. learning) as opposed to short-term memorization [63].

9.2.2 Evaluating the Impact of Real-Time Feedback on Performance

In addition to evaluating learning and training transfer with the above procedure, Study MRIPS-Learning also evaluated the impact of real-time feedback on learner performance. To evaluate the impact of real-time feedback on performance, we compared cognitive, psychomotor, and affective performance of Study MRIPS-Learning participants to that of participants in Study MRIPSx2 (Section 4.3). The content of the MRIPS interactions in Study MRIPSx2 and Study MRIPS-Learning is identical (i.e. same virtual human appearance, behaviors, responses, and critical moments). Other than real-time feedback, the difference between the two studies was the use of the Wizard-of-Oz in Study MRIPSx2. In Study MRIPSx2, if a participant repeatedly asked a question (corresponding to a response in the virtual human response database) but did not get a response due to speech recognition or speech understanding failure, the experimenter triggered the virtual human to provide the appropriate response. This mechanism was not used in Study MRIPS-Learning. For MRIPS-CBE to function in its intended role as an on-demand learning opportunity, responses to participant speech depended wholly on speech recognition and understanding modules. This difference has the potential to put Study MRIPS-Learning participants at a comparative
disadvantage in tasks involving speech (e.g. taking a complete medical history). In the sections detailing this analysis, we discuss the performance of the speech interface and how measures were designed to mitigate the effect of this difference.

9.2.3 Control Groups for Investigating the Validity of Study Results

Prior to conducting Study MRIPS-Learning, we conducted a pilot study which had two goals: determining the impact of a single MRIPS interaction on performance in CBE of an SP, and determining the impact of a single SP interaction on performance in subsequent SP interactions. This pilot study provided baseline data which allowed us to evaluate the validity of the Study MRIPS-Learning results.

The pilot study consisted of two phases. In the first phase, Group MRIPS-SP, eleven medical students with the same experience level as Study MRIPS-Learning participants, performed an MRIPS interaction. One week later, this was followed by an SP interaction. The MRIPS interaction provided real-time feedback of coverage and correct palpation pressure, and provided post-experiential feedback of coverage, pressure, visual inspection, and breast history completeness in a web-based interface. This web-based interface also provided self-driven feedback of affective performance in the form of reviewable videos of the participant’s and an expert’s interactions. Also in phase one, Group SP, an additional eight medical students, completed only the SP interaction. Medical educators, clinicians, and residents rated participant performance in the SP interactions of both groups by reviewing video of the interactions. Comparing the two groups’ performance in the SP interaction allows us to evaluate how much of an impact one MRIPS interaction has on performance in a CBE of an SP.

The second phase involved five participants from Group MRIPS-SP performing a CBE of an SP in the women’s health clinic. These participants were evaluated in-
person by a single expert. In clinic, exams can not be videotaped for review by multiple experts. This SP interaction came approximately one month after Group MRIPS-SP’s first SP interaction. These two SP interactions mimic the SP Pre-test and SP Post-test interactions of Study MRIPS-Learning. Evaluating improvement in performance between these two SP interactions allows us to determine whether the SP Pre-test of Study MRIPS could cause observed improvement from SP Pre-test to SP Post-test. If, in this pilot study, there is no improvement (or there is a decrease in performance) between the two SP interactions, than any improvement from SP Pre-test to SP Post-test must be due to the three MRIPS-CBE interactions. Analysis of performance in this pilot study and its impact on the validity of the Study MRIPS-Learning results is presented in Section 9.8.

9.3 Population

Participants were recruited from 3rd-year medical students beginning a women’s health clerkship at the Medical College of Georgia. The total size of this population was 30 students. Prior to conducting the study, we conducted a power and sample size analysis using G*Power3 [146]. With 5 repetitions and an alpha of 0.05, we could expect to find large effects (Cohen’s $f^2 >= 0.4$) with a sample size of 17 participants.

Our collaborators at the Medical College of Georgia were able to recruit 16 of the 30 medical students for the study. Participants were unpaid volunteers. With a study taking place over the course of one month, we expected to lose a few participants. All 16 participants completed the background portion of the study in which they received a lecture in which an expert clinician demonstrated a CBE. These participants also completed the background survey. Of these participants, only 12 participants completed the SP Pre-test and first MRIPS-CBE interaction, and only 8 participants
completed the entire study. With 8 participants, we can still find large effects ($f^2 = 0.5$). Statistically, these effects must be larger than those at $f^2 = 0.4$ (i.e. the test statistic, e.g. $t$ or $F$, must be larger to indicate significance).

Because the population was small, we view Study MRIPS-Learning as only a first step at investigating learning in MRIPS. For evaluation of learning we analyze the data of the 8 participants who completed the entire treatment. In evaluating the impact of feedback, we will also incorporate the data of the 12 participants who completed the SP Pre-test and first MRIPS interaction. Larger historical control groups are also used in comparison to evaluate the impact of feedback.

All participants had the same background in clinical breast examination. None of the participants had performed a CBE of an SP or a real patient. Participants’ experience in CBE was limited to receiving a lecture and demonstration from an expert clinician. During the course of the study, no participants received education or practice in clinical breast examination from any sources other than the study’s treatments.

9.4 Statistical Analysis

Because a small population was used, we can not assume a normal distribution of participant performance. For this reason we primarily used non-parametric tests in our analysis.

When comparing changes in performance from SP Pre-test to SP Post-test, we used the Wilcoxon signed-rank test [147]. The Wilcoxon test is a non-parametric version of the paired samples t-test, used to evaluate within-subjects change in a 2-level repeated measure (i.e. before and after treatment).
To compare changes in performance throughout the three MRIPS treatments, the Friedman test was used. The Friedman test is a non-parametric version of a one-way analysis of variance with a k-level repeated measure (k > 2) [148].

For dichotomous (binary outcome) categorical variables, McNemar’s test is used to assess significant change in a 2-level repeated measure (i.e. before and after treatment) [149].

When comparing to historical controls, if the control group is significantly large enough to assume a normal distribution, we used 2-way analysis of variance (ANOVA) because of the relative common usage and familiarity to readers. For all tests, the acceptable significance level is set at $\alpha = 0.05$, giving us a minimum confidence level of 95% in rejecting a null hypothesis. Since comparison to historical controls is a between-subjects comparison, we also use Fisher’s exact test when comparing dichotomous variables between-subjects. Fisher’s exact test is a non-parametric substitute for the Chi-square test which is used with small samples sizes (i.e. when the expected value for any cell in the 2x2 contingency table is less than 10) [150].

When describing changes in performance, we will occasionally refer to the median change. In the small population of Study MRIPS-Learning, a single outlier could significantly impact the mean, but the median is less sensitive to these outliers.

### 9.5 Cognitive Performance

The cognitive tasks on which participants were evaluated are:

Breast history completeness. Evaluation of the patient’s medical history and risk factors for breast cancer. The participant must recall the important questions to ask in order to evaluate the patient’s history of present illness (breast pain), risk factors, and relevant medical history.
Visual inspection completeness. Performing a complete visual inspection. The participant must recall and have the patient assume the three poses required for a complete visual inspection: arms relaxed at sides, hands on hips with chest flexed, and arms raised above head.

9.5.1 Measures

9.5.1.1 Breast history completeness

The completeness of the breast history is measured as the number of questions asked from a 27 item list (Table 9-1). These items are taken from validated instruments used to evaluate students in breast history taking of SPs and real patients at the University of Florida’s College of Medicine and the Medical College of Georgia.

Twenty-one of these items are present in the procedural checklist real-time feedback described in Chapter 7. During the MRIPS-CBE interactions, these items are displayed on the screen above the virtual human, ordered by topic (history of present illness, medical history, family history, social history). This feedback is expected to improve performance in the cognitive task of recalling the critical questions to ask and aid in keeping track of the items previously queried.

All twenty-seven items are reviewed in the post-experiential feedback viewed by participants after each MRIPS-CBE interaction. By leaving six items off of the real-time feedback and displaying them only in post-experiential feedback, we are able to observe whether real-time feedback has benefits beyond those of post-experiential feedback for the breast history taking task.

The MRIPS-CBE speech interface automatically records participants’ asking of items on this list. For the SP interactions, the completeness was also assessed by reviewing video.
The experimenter also reviewed video of participant performances to confirm or correct the automated rating. A discussion of the automated rating is provided in Section 9.5.5. The manual review was performed to remove two types of errors: false positives and false negatives. False positives occurred when the participant asked a question and received an unrelated response which provided information related to an item in the list. Eight false positives occurred across 28 MRIPS interactions. It is possible that receiving this unrelated information would cause the participant to not query this item further. However, in practice, we observed that participants typically pursued their original line of questioning until they received a related answer, and then asked for confirmation of the earlier unrelated piece of information. Even with this behavior, because we did not include false positives in participants’ MRIPS scores, these scores may be viewed as a “lower bound” on actual performance. In contrast, false positives were included in SP scores. With the SP, false positives occurred when SPs volunteered more information than the script called for in response to a participant question. There were two instances (out of 20 SP interactions) in which a false positive was not counted. In both instances, the SP volunteered information that her mother had breast cancer in a critical moment speech at the end of an interaction (after the exam). The participants had not asked about family history and were closing the patient-doctor interview, when the SP used the speech (similar to) “I’m scared it could be cancer because my mom died of breast cancer.” The choices made in how to handle scoring false positives reflect the qualities of the virtual human and real human interactions: the virtual human interaction follows a more rigid question-answer form than the more open-ended conversation with the SP.
False negatives occurred when participants repeated a question multiple times but could not get a related response. These occurred due to consistent speech recognition and understanding errors specific to that question (due to southern accent, poor enunciation, or odd phrasing which could not be matched to the correct response). There were 43 false negatives across the 28 MRIPS interactions, an average of ~1.5 per interaction and a mode of 2 per interaction. SP interactions did not have false negatives.

9.5.1.2 Visual inspection completeness

In the MRIPS interactions, the completeness of the visual inspection was evaluated automatically. For the SP interactions, the completeness of the visual inspection was evaluated by review of video. For a complete visual inspection, the patient should be examined with arms relaxed by her sides, hands pressed on hips, and arms raised above head. Participants were judged on whether they performed any visual inspection (one or more poses) and whether they performed a complete visual inspection (all three poses). Participants’ visual inspections are guided in MRIPS, as the procedural checklist real-time feedback in MRIPS displays icons depicting the poses used in visual inspection (Figure 7-2).

9.5.2 Hypotheses

- **Hypothesis Breast History Completeness Learning and Transfer**: The completeness of participants’ history taking and breast cancer risk assessment will increase from SP Pre-test through the three MRIPS interactions. The completeness of participants’ history taking and breast cancer risk assessment will increase significantly from SP Pre-test to SP Post-test.
  - **Null hypothesis**: Participants’ scores on the breast history completeness instrument (Table 9-1) will not significantly improve throughout the two SP and three MRIPS interactions.
• **Hypothesis Breast History Completeness Feedback Impact**: For the breast history items which are included in the real-time feedback, participants’ completeness in the first MRIPS interaction will be significantly greater than in historical data in which real-time feedback was not provided.
  
  o **Null hypothesis**: Completeness of participants’ breast history taking will not be significantly different from historical MRIPS data.

• **Hypothesis Visual Inspection Learning and Transfer**: The number of participants who perform a complete visual inspection will increase from SP Pre-test through the three MRIPS interactions. The number of participants performing a complete visual inspection will increase significantly from SP Pre-test to SP Post-test.
  
  o **Null hypothesis**: The number of participants performing a complete visual inspection will not significantly improve throughout the two SP and three MRIPS interactions.

A summary of results of acceptance and rejection of these hypotheses is shown in Table 9-2.

**9.5.3 Results: Breast History Learning and Training Transfer**

Means and standard deviations as well as ranges are shown in Table 9-3. Scores for the eight participants are visualized in Figure 9-2.

Participants significantly improved the completeness of their breast histories from SP Pre-test to SP Post-test, by Wilcoxon test: $Z = 2.4$, $p = 0.02$. Seven of eight participants increased the completeness of their breast histories and one participant queried the same number of items (though not the same items) in pre-test and post-test. The median change was an increase by 5.5 items (20% of the 27 items).

The largest component of the improvement was from SP Pre-test to MRIPS #1, as all participants improved their completeness by an average of 8.3 items. Completeness in MRIPS #1 was significantly greater than in the SP Pre-test, by a Wilcoxon test: $Z = 2.5$, $p < 0.01$. 


Throughout the three MRIPS interactions, participants did not significantly increase the number of breast history items asked on average. From MRIPS #1 to MRIPS #2, three participants improved, two scored the same, and three had their scores decrease. From MRIPS #2 to MRIPS #3, five increased their scores, and three had their scores decrease. Overall, from MRIPS #1 to MRIPS #3, the average score increased by one question asked. For six of the participants, the real-time and post-experiential feedback regarding the completeness of their breast history appears to have caused a ceiling effect in the first or second MRIPS interaction. The other two participants (participants 20 and 22 in Figure 9-2) improved in each repeated MRIPS interaction. These two participants represented the lower bound on performance in each of MRIPS #1 and MRIPS #2 interactions. The effect of the repetition was to bring these participants up to the level of performance achieved by the other participants.

With respect to the ceiling effect experienced by the majority of participants, the ceiling was set at or close to 21, the number of items appearing in the real-time feedback. For measuring the “ceiling” value, we included false positives, as false positives caused the items to appear “checked off” in the feedback. As such, participants could not use the feedback to keep track of whether they had actually asked about the false positive item. In MRIPS #1 and MRIPS #2, six of 8 participants asked 19 or more of the 21 items appearing in the real-time feedback. In MRIPS #3, all participants asked 20 or 21 items. Reasons for not asking all 21 items may include not wanting to break up the topic flow (i.e. if an item related to medical history is skipped and the participant moves on to family history, the participant may not feel that it is necessary to return to medical history later), frustration with speech recognition (which
was apparent only in interaction by one participant), or not wanting to rely completely on the real-time feedback.

Some participants clearly did rely on the real-time feedback, as the real-time feedback guided their line of questioning. Typically, medical students have a difficult time asking all questions within a specific topic (e.g. family history) before moving along to the next topic [84]. The real-time feedback appeared to help guide the progression of questions and topics, as 5 of 8 participants followed the progression displayed in the real-time feedback exactly for at least one MRIPS interaction. One participant followed the progression exactly for all MRIPS interactions, with appropriate introductions to each topic. This participant followed the feedback exactly but was not progressing mindlessly through the items displayed by the feedback, i.e. trying to game the system by “checking off” all items.

If a participant asked a nonsensical question for the purpose of “checking off” an item without displaying knowledge of the medical information the item refers to, we did not count that item in the participant’s total. We observed only two instances of this: in MRIPS #1 and MRIPS #2 interactions, the same participant asked “are you having any problems with your hormones” in reference to the “hormones” item. The hormones item refers to the risk factor of taking replacement hormones (if the patient is post-menopausal) or taking birth control. By the third MRIPS interaction, this participant realized what “hormones” referred to and asked an appropriate question about hormone use.
Aside from the five participants following the progression exactly, two additional participants followed the progression of topics, but asked questions within each topic in an order different from the progression shown in the feedback.

It appears clear that the feedback and repetition helped participants learn which important questions to ask the patient. From SP Pre-test to MRIPS #3, the median improvement was 9.5 items or 35% of the list. Most of this improvement was retained with the removal of the feedback in the SP Post-test. From MRIPS #3 to SP Post-test, the median retention rate was 84% (calculated by dividing SP Post-test performance by MRIPS #3 performance, and taking the median result among participants).

The improvement in number of items asked does not appear to be a random effect of guessing at possibly important questions, as there were specific items which a significantly larger percentage of the population queried in SP Post-test than in SP Pre-test. In the SP Pre-test, no participants asked about the age at which the patient began menarche. Early onset of menarche is a well-confirmed correlate of increased risk for breast cancer [151]. All participants asked about onset of menarche in MRIPS #3, and 5 of 8 remembered to ask in the SP Post-test. Other such items are summarized in Table 9-4. Of the items that received large (+3 or more) increases, the patient’s age is the only item not included in the real-time feedback. The patient age item is part of the post-experiential feedback. A discussion of the comparative impact of the real-time and post-experiential feedback is given in Section 9.5.4.

As the result of this analysis, we accept Hypothesis Breast History Completeness Learning and Transfer. Participants significantly increased the number of breast history items asked from before to after the MRIPS treatment. The improvement was not solely
due to repetition, but due to repetition with feedback, as a ceiling effect was observed starting in the first MRIPS interaction. The improvement was not from random guesswork, as specific critical items saw significant improvement. MRIPS’ combination of repetition while receiving real-time feedback led to significant improvement in the completeness of learners’ breast histories in the real-world scenario of a CBE of an SP.

9.5.4 Results: Impact of Feedback on Breast History Completeness

To evaluate whether the procedural checklist real-time feedback improved breast history completeness in MRIPS, we compared participants in Study MRIPS-Learning to a historical control group. The historical control group consisted of experienced (having performed >= 5 CBEs) and inexperienced (0-4 CBEs) students from Study MRIPSx2 (Section 4.3). This group contained 29 participants. The residents and clinicians in Study MRIPSx2 were not included because, as discussed in Section 4.3, the number of questions asked is not a valid means of querying expert performance. To gain a slight increase in power for this comparison, we included all 12 participants in Study MRIPS-Learning who completed the first MRIPS interaction. The addition of the four participants that only completed one MRIPS interaction actually served to decrease the mean score for participants in Study MRIPS-Learning.

The instrument used to evaluate breast history completeness in the historical control contained 20 items. These items were a subset of the 21 items in the real-time feedback of Study MRIPS-Learning (non-highlighted items in Table 9-1). All participants in Study MRIPS-Learning queried the one additional item that was not present in Study MRIPSx2, thus their scores were reduced by one point in order to compare with the Study MRIPSx2 scores.
Because the groups were large enough to assume a normal distribution (and from a histogram, a normal distribution appeared to be an acceptable fit), a univariate analysis of variance was conducted.

Participants who viewed the procedural checklist in real-time performed significantly more complete breast histories than the historical control group which did not receive the real-time procedural checklist feedback. Study MRIPS-Learning participants outperformed the Study MRIPSx2 participants by an average of nearly 7 items: Study MRIPS-Learning = 17.3 +/- 4.3; Study MRIPSx2 = 10.6 +/- 3.0. This was significant by ANOVA: \( F(\text{d.f.} = 1, n = 41) = 33.2, p < 0.001. \)

From this result we accept Hypothesis Breast History Completeness Feedback Impact. The real-time feedback of the procedural checklist improved performance in breast history completeness over that observed in the historical control group.

This result shows that real-time feedback is more effective than no feedback. By investigating the data from the 8 participants who performed three MRIPS interactions, we find that real-time feedback in addition to post-experiential feedback has benefits over post-experiential feedback alone.

In all of the SP and MRIPS interactions, participants asked a higher percentage of the items in the real-time feedback than the items reviewed only in the post-experiential feedback. This difference in percentages increased with the repetition as shown in Figure 9-3. For each interaction, the difference is significant by Wilcoxon test, at \( p = 0.05 \) for SP Pre-test and at \( p < 0.01 \) for the remaining interactions. Both categories of items were asked more often in the SP Post-test than in the SP Pre-test, but the improvement in items appearing in the real-time feedback is significant and the
improvement in post-experiential items was not significant by Wilcoxon test:
improvement in real-time items: $Z = 2.5, p < 0.01$; improvement in post-experiential
items $Z = 0.9, p = 0.5$, ns. Participant 21 may have benefitted from the post-experiential
feedback more than the other participants (see Figure 9-2) as this participant’s
improvement from MRIPS #3 to SP Post-test was due to asking items in the post-
experiential feedback. Interestingly, this participant did not spend more time with the
post-experiential feedback than other participants. Participant 21 spent ~80 sec. with
the post-experiential feedback which was similar to the average time of ~75 sec
(ranging from ~30 sec. to ~102 sec.).

From these observations, it appears as though there is benefit from practicing with
a visible list of topics to query, along with the ability to cross off the items. One might
wonder why a graphical display is needed for this task; would a piece of paper be
adequate. Students are not given a list of items during SP or real patient interviews. In
the practice scenario of MRIPS, visualizing this information on the screen above the
virtual human’s head has the advantage of not incurring an increased cognitive load
from looking back and forth between the screen and a piece of paper.

9.5.5 Visual Inspection Learning and Training Transfer

To evaluate improvement in visual inspection of the patient’s breasts, we analyzed
how many participants performed any visual inspection in each interaction as well as
how many of these visual inspections were complete. An inspection was complete if it
involved the three poses of arms relaxed, hands on hips, and arms raised above head.

From SP Pre-test to SP Post-test, there was a trend towards significant
improvement in the number of participants performing visual inspections. In the SP Pre-
test, two of the eight participants performed a visual inspection. This number increased
to six of eight participants in the SP Post-test. This trended towards significance at $p = 0.06$ from McNemar’s test.

From SP Pre-test to SP Post-test, there was a significant increase in the number of participants who performed complete visual inspections. No participant performed a complete visual inspection in the SP Pre-test, while five of eight performed complete inspections in the SP Post-test. This improvement was significant by McNemars at $p = 0.03$.

The pattern of improvement in visual inspection mirrored that of improvement in breast history completeness, as shown in Table 9-5. The improvement from SP Pre-test to MRIPS #1 in the number of participants performing any visual inspection trended towards improvement ($p = 0.063$); the improvement in the number of participants performing a complete visual inspection was significant ($p = 0.03$). Among the three MRIPS interactions, there were no significant changes in the number of participants performing a visual inspection, nor were there significant changes in the number of participants performing a complete visual inspection.

It was expected that in all MRIPS interactions, all participants would perform a complete visual inspection, as the three poses were included in the real-time procedural checklist feedback. One participant did not perform a visual inspection in any of the MRIPS or SP interactions. For this participant, not performing any visual inspection appeared to be due to participant preference, as there were no technical difficulties experienced in this task in MRIPS. However, the experimenter did not ask the participant if this was the case.
As the completeness of participants’ visual inspections increased significantly from the SP Pre-test to MRIPS #1 and from the SP Pre-test to the SP Post-test, we accept Hypothesis Visual Inspection Learning and Transfer. MRIPS provided repetitive practice which reinforced the need to perform a visual inspection and the three poses for a complete visual inspection. This knowledge transferred to the SP Post-test, indicating learning and training transfer from practice with MRIPS.

9.5.6 Discussion

Repetitive practice with MRIPS in which the procedural checklist feedback was provided led to improved performance in completeness of both breast history taking and visual inspection. For breast history completeness, real-time feedback led to improved performance in comparison to performance when no feedback is provided. Real-time feedback also appears more effective at helping learners retain information than post-experiential feedback. Providing real-time feedback for these cognitive tasks requires only a speech interface (i.e. does not require a touch interface), thus MRIPS is not unique among interpersonal simulations in being able to provide this feedback. However, MRIPS is the first interpersonal simulation to provide real-time feedback of this type and evaluate its impact. The combination of the on-demand (repetitive) learning opportunity and real-time feedback provided by MRIPS led to improvement in cognitive tasks in a real-world interpersonal scenario.

9.6 Psychomotor and Cognitive-Psychomotor Performance

In our analysis, we grouped the psychomotor and compound cognitive-psychomotor tasks together. The psychomotor and cognitive-psychomotor tasks on which participants were evaluated are:

- Palpating all breast tissue, regardless of pressure.
Palpating the entire breast, i.e. complete coverage
- Palpating the axillary, supraclavicular, and infraclavicular areas.
- Using correct (deep) pressure to palpate the entire breast.
- Palpating along the correct (expert) pattern-of-search.
- Determining whether palpated breast tissue is normal or a mass, i.e. finding actual masses and not reporting false positive masses.

9.6.1 Measures

9.6.1.1 Coverage and correct pressure

MRIPS-CBE is able to automatically calculate the percent area of the cone of the breast that is palpated at each of the levels of pressure (light, medium, high, too-high). As discussed in Section 7.4, this requires a calibration step in which an expert performs an exam using MRIPS. For Study MRIPS-Learning, the calibration exam was performed by a clinician who was considered to be the breast examination expert at the Medical College of Georgia.

Although MRIPS calculates percent area at light, medium, high, and too-high levels of pressure, these were simplified to three measures for recording learner performance. These three levels were percent area palpated with superficial pressure (light, medium), high pressure, and too-high pressure. An ideal performance would palpate 100% of the breast at high pressure.

However, it was not clear how to integrate superficial and too-high measures into a rating of use of correct pressure. For example, is it better to palpate at 0%-70%-20% or 5%-70%-15%? The first palpates 90% at deep pressure, while the second decreases the use of too-high pressure at the expense of palpating only 85% at deep pressure. We define deep pressure as high or greater pressure (high or too-high pressure). To determine how to integrate superficial and too-high measures into an overall score of performance, we consulted with two medical experts in CBE. Both experts expressed that it was better to palpate "too-hard" than to not palpate deep enough, e.g. “the patient
would rather hear you say sorry for pressing too hard than sorry you missed a lesion because you did not [press hard enough]” (Carla M. Pugh, personal communication). Based on this feedback from experts, we used two measures to evaluate coverage and use of correct pressure.

1. Coverage: The percent area palpated at light or greater pressure is used as the measure of coverage.
2. Correct pressure: The percent area palpated at high or too-high pressure is used as the measure of correct deep pressure.

An additional question lies in determining what constitutes a significant change in percentage of breast tissue palpated. This significance differs from significance in a statistical sense; here a significant change refers to the smallest percentage change that matters to the outcomes of the exam. As one important outcome is finding masses, we use the size of the masses present in the breast to calculate a significant percentage. The breast used in MRIPS is approximately 7” (17.8 cm) long by 5” (12.7 cm) wide. The masses used were approximately 2 cm x 2 cm. Thus a mass represents 1.8% or roughly 2% of the area of the breast. Not palpating 2% of the breast area could result in not finding a mass, thus we chose 2% as a threshold of significance when discussing changes in participant performance in coverage and in palpating the breast at deep pressure. If a participant can improve by 2%, this hypothetically increases the participant’s ability to find masses present in the breast. In determining what constituted a “passing” score in area palpated, our medical collaborators expressed that they would accept ≥90% as indicating competency. However, there exists no current standard for competency in percent of breast tissue palpated [9].

In the SP interactions, coverage and use of correct pressure can not be measured with the same precision as in MRIPS-CBE. Although the SP wore a silicone breast vest
(to incorporate breast masses), our method of quantitatively measuring coverage and pressure (Section 7.4) could not be applied due to an inability to have the SP lie in the same position for each exam. This would be required to maintain a calibration between the camera which tracked palpation position and the sensors placed in the vest. Additionally, the breast vest could not accommodate sensors, as it did not provide a rigid surface beneath the silicone on which to seat the sensors. The sensors must be placed on a rigid surface in order to maintain a consistent relationship between force applied and value reported. Adding a rigid surface would make the vest uncomfortable for the female SPs.

Because of these limitations, to evaluate psychomotor performance in the SP interaction, five experienced medical professionals and two medical students rated performance from reviewing video of participants’ exams. Three MDs, one RN, one resident, and two 3rd-year medical students performed the video rating. The instrument used to rate psychomotor performance was a validated instrument used to evaluate students in the women’s health clerkship at the Medical College of Georgia. All video raters were trained in using the video rating instrument to reduce variability between the five experienced raters and the two inexperienced (medical student) raters. Items from the video rating instrument pertaining to coverage and pressure are listed in Table 9-6.

Palpation of the ancillary areas of tissue included in the CBE was evaluated separately. Palpation of the axillary, supraclavicular, and infraclavicular areas was evaluated as three dichotomous (yes or no) variables. MRIPS-CBE reported palpation of these three areas automatically. For the SP interactions, the video raters determined
if these areas were palpated. Participants’ performance in this task is rated as the
number of areas palpated, from 0 to 3.

9.6.1.2 Correct pattern of search

In evaluating the participant’s pattern-of-search, MRIPS-CBE calculated a total
deviation of the learner’s pattern from the expert’s pattern. This was calculated by
Equation 9-1, a summation over all segments in the learner’s pattern, in which \( s_i \) is the
\( i^{th} \) segment of the learner’s pattern, \( v_i \) is the matching segment of the expert’s pattern,
and \( l_i \) is the length of the learner’s \( i^{th} \) segment. The total deviation is the most sensitive
measure calculated by MRIPS-CBE of the difference between learner pattern and
expert pattern, but total deviation does not account for the overall length of the pattern.
If two learners make one large mistake (one segment with high deviation) both are
penalized equally, even if this mistake represents a small percentage of one learner’s
pattern and a large percentage of the other learner’s pattern. To account for overall
pattern length, the total deviation is normalized by the total pattern length, as in
Equation 9-2. This measure is the normalized deviation, and is used to in Study
MRIPS-Learning as the measure of correctness of participants’ patterns-of-search in
MRIPS-CBE.

\[
\sum_{i} \left( l_i \times \arccos \left( \frac{\bar{s}_i \cdot \bar{v}_i}{90^\circ} \right) \right) \quad (9-1)
\]

\[
\left( \sum_{i} \left( l_i \times \arccos \left( \frac{\bar{s}_i \cdot \bar{v}_i}{90^\circ} \right) \right) \times 90^\circ \right) / \sum_{i} l_i \quad (9-2)
\]

For the SP interactions, pattern of search can not be evaluated quantitatively or
with the same precision as in MRIPS. To evaluate participants’ pattern of search,
exerts reviewed video to determine whether participants adequately used the vertical
strip pattern, a different systematic pattern (e.g. spiral, spokes), or did not use a systematic pattern.

9.6.1.3 Finding masses

In both MRIPS and SP interactions, participants reported the number, location, and quality (fixed or mobile, hard or rubbery) of masses found. This was reported in a note written by participants after the interaction. Participants drew a diagram of the breast which indicated the location of each mass found.

The SP breast contained one hard, fixed mass at the base of the breast (Figure 9-4, B) which was placed at the chest wall. The MRIPS breast contained a similar hard, fixed mass at the chest wall, placed a few centimeters medially from the nipple (Figure 9-4, A). This mass was judged by clinicians to be of similar difficulty to find as the mass in the SP breast. The MRIPS breast also contained a second mass at the tip of the breast cone. This mass was soft and semi-mobile, i.e. it could be moved 1-2 cm but remained in the same area for all participants (Figure 9-4, A). All masses were roughly the same size. Masses were not spherical but were generally “round,” convex shapes and the long axis of each mass was ~2 cm in length. The experimenter and experts reviewed the notes to determine which reported masses were actual masses and which were false positives.

In the SP interactions, participants were graded on whether they found the actual mass and the number of false positives reported. In the MRIPS interactions, participants were graded on the number of actual masses found and false positives reported.
9.6.2 Hypotheses

- Hypothesis Coverage Learning: Participants’ coverage will increase with repeated MRIPS interactions.
  - Null hypothesis: The percent of breast area palpated at light or higher pressure will not significantly increase with repeated MRIPS interactions.
- Hypothesis Pressure Learning: Participants’ use of correct pressure will increase with repeated MRIPS interactions.
  - Null hypothesis: The percent of breast area palpated at deep or higher pressure will not significantly increase with repeated MRIPS interactions.
- Hypothesis Coverage Transfer: The coverage of participants’ exams will improve from SP Pre-test to SP Post-test.
  - Null hypothesis: Expert ratings of participants’ coverage will not significantly increase from SP Pre-test to SP Post-test.
- Hypothesis Pressure Transfer: The pressure of participants’ exams will improve from SP Pre-test to SP Post-test.
  - Null hypothesis: Expert ratings of participants’ use of correct pressure will not significantly increase from SP Pre-test to SP Post-test.
- Hypothesis Coverage Feedback Impact: The presence of real-time feedback of coverage will result in improved coverage in participants’ exams in MRIPS.
  - Null hypothesis: Participants coverage in MRIPS #1 will not be significantly more complete than historical coverage data from Study MRIPSx2.
- Hypothesis Pressure Feedback Impact: The presence of real-time feedback of use of correct pressure will result in improved use of deep pressure in participants’ exams in MRIPS.
  - Null hypothesis: The area palpated at deep pressure by participants in MRIPS #1 will not be significantly greater than area palpated at deep pressure in historical data from Study MRIPSx2.
- Hypothesis Pattern-of-Search Learning: Deviation of participants’ pattern-of-search will decrease with repeated MRIPS interactions.
  - Null hypothesis: Participants’ normalized deviation from the expert pattern-of-search will not decrease significantly with repeated MRIPS interactions.
- Hypothesis Pattern-of-Search Transfer: Expert ratings of the correctness of participants’ pattern-of-search will improve from SP Pre-test to SP Post-test.
  - Null hypothesis: Expert ratings of the correctness of participants’ pattern-of-search will not significantly improve from SP Pre-test to SP Post-test.
- Hypothesis Pattern-of-Search Feedback Impact: The presence of real-time feedback of the correctness of participants’ pattern-of-search will result in a significantly larger percent of participants using the vertical strip pattern than in historical data from Study MRIPSx2.
  - Null hypothesis: The percentage of participants using the vertical strip pattern in MRIPS #1 will be significantly larger than the percentage of participants using the vertical strip pattern in Study MRIPSx2.
- Hypothesis Finding Masses Learning: The number of masses participants find will increase with repeated MRIPS interactions.
o Null hypothesis: The number of masses found will not significantly increase throughout the MRIPS interactions.

- Hypothesis Finding Masses Transfer: The number of participants finding the mass in the SP breast will increase from SP Pre-test to SP Post-test.
  o Null hypothesis: The number of participants finding masses will not increase significantly from SP Pre-test to SP Post-test.

- Hypothesis False Positive Masses Learning: The number of false positive masses reported by participants will decrease with repeated MRIPS interactions.
  o Null hypothesis: The number of false positive masses reported by participants will not significantly decrease with repeated MRIPS interactions.

- Hypothesis False Positive Masses Transfer: The number of false positive masses reported by participants will decrease from SP Pre-test to SP Post-test.
  o Null hypothesis: The number of false positive masses reported by participants will not significantly decrease from SP Pre-test to SP Post-test.

A summary of results of acceptance and rejection of these hypotheses is shown in Table 9-7.

9.6.3 Results: Coverage and Pressure Learning and Transfer

9.6.3.1 Coverage learning

Participants’ performance in palpating the three ancillary areas (supraclavicular, infraclavicular, and axilla) is listed in Appendix D.2. In SP Pre-test, only three of eight participants palpated one or more of these areas. The average number of areas palpated was 0.6 ± 0.9. Six of eight participants improved by palpating more of these areas in MRIPS #1 than in the SP Pre-test; one participant palpated one less area and one participant palpated the same number of areas. This improvement is significant by Wilcoxon test: $Z = 1.9, p = 0.047$. This suggests that the real-time feedback provided by MRIPS assisted participants in palpating these areas. From MRIPS #1 to MRIPS #3, five of eight participants improved, two did not change, and one decreased in performance. This improvement was not significant (Wilcoxon test: $Z = 0.96, p = 0.17$). In MRIPS #1, participants palpated an average of 1.5 ± 1.1 of these areas; this increased in MRIPS #3 to an average of 2.1 ± 1.1 of these areas. From SP Pre-test to
MRIPS #3, six of eight participants improved and two had no change. This improvement was significant by Wilcoxon test, $Z = 2.3, p = 0.16$. Most importantly, the number of participants palpating all three areas increased from zero to four, and the number of participants palpating at least one area more than doubled, from three to seven participants. For the task of palpating the tissue surrounding the breast, practice with MRIPS appears to result in learning.

Palpation of the breast was measured with different mechanisms and precisions in MRIPS and SP, so we can not compare SP Pre-test and MRIPS #1 to investigate learning; instead we are only able to look for improvement throughout the three MRIPS interactions.

We observed a ceiling effect occurring in MRIPS #1 for the majority of participants. Five of eight participants palpated 90% or more of the breast tissue in MRIPS #1. This is a significantly higher proportion than the 5 of 57 participants palpating at 90% or more in the historical control group of MRIPSx2 (by Fisher’s exact test, $p < 0.005$). This suggests that the real-time feedback of coverage resulted in a ceiling effect in MRIPS; further discussion of the impact of feedback on performance is given in Section 9.6.4.

Because of their high level of performance in MRIPS #1, most participants decreased in coverage from MRIPS #1 to MRIPS #2. Using our 2% threshold for significant change, 5 of 8 participants decreased from MRIPS #1 to MRIPS #2. All 5 had palpated at >90% in MRIPS #1. Three of the participants whose performance decreased remained at or above 90%. The three participants improving from MRIPS #1 to MRIPS #2 improved from <80% to >89%.
MRIPS #3 data was not available for one participant due to an equipment error. From MRIPS #2 to MRIPS #3, four of seven participants increased significantly, one increased by <2%, and two decreased. One of the participants who decreased palpated above 90% in both MRIPS #2 and MRIPS #3.

From participants’ first to last MRIPS interaction, three participants improved significantly, three participants changed by less than 2%, and two participants’ performance decreased (including the one participant who only had data for MRIPS #1 and MRIPS #2). Overall, participants performed well in this task, as in MRIPS #3, five of seven participants palpated >90% of the breast and another participant palpated >89%. However, due to the ceiling effect, there was not a statistically significant improvement from MRIPS #1 to MRIPS #3, so we must reject Hypothesis Coverage Learning. We do so noting that performance in palpating the three areas of tissue surrounding the breast improved significantly, and performance in palpation of the breast was significantly higher than performance observed in past MRIPS interactions. A larger population and increased number of repetitions with MRIPS is expected to provide clearer evidence of learning within MRIPS.

9.6.3.2 Coverage transfer

From SP Pre-test to SP Post-test, participants significantly increased their coverage, palpating more of the breast tissue. In SP Pre-test, 2 of 8 participants were rated as completely palpating the cone of the breast only, one participant was rated as completely palpating the entire breast, and 5 of 8 participants were rated as incompletely palpating the cone of the breast. In the SP Post-test, 4 of 8 participants were rated as completely palpating the cone of the breast and 4 of 8 participants were rated as completely palpating the entire breast. This represented an improvement for 6
participants and no change for 2 participants. These changes represented significant improvement, by a Wilcoxon test: $Z = 2.2, p = 0.02$.

In palpating the ancillary areas of tissue included in the breast exam (supraclavicular, infraclavicular, and axilla), participants trended towards significant improvement from SP Pre-test to SP Post-test. In the SP Pre-test, 2 of 8 participants palpated two of these areas, and one participant palpated one of these areas. The other five participants did not palpate any of these three areas. In the SP Post-test, four participants palpated two of the three areas, three participants palpated one of the three areas, and one participant did not palpate any of the three areas. This represented an improvement for five participants, no change for one participant, and a decrease in performance for one participant. These changes trended towards significant improvement, by a Wilcoxon test: $Z = 1.7, p = 0.08$. From this trend in improvement in the coverage of ancillary areas of tissue and the significant improvement in ratings of coverage of the breast tissue, we accept Hypothesis Coverage Transfer.

9.6.3.3 Pressure learning

As with coverage, palpation pressure was measured using different mechanisms and precisions in MRIPS and SP, so we can not compare SP Pre-test and MRIPS #1 to investigate learning; instead we are only able to look for improvement throughout the three MRIPS interactions. Performance in the three MRIPS interactions is shown in Figure 9-5.

Participants performed well in MRIPS #1, with 7 of 8 participants (88%) palpating more than 60% of the breast with deep pressure. For comparison, in Study MRIPSx2, only 58% (33 of 57 participants) palpated >60% of the breast with deep pressure. From MRIPS #1 to MRIPS #2, four participants improved, one participant had no significant
change (<2%), and for three participants the percent area palpated decreased. One of the participants who performed worse in MRIPS #2 palpated 95% of the breast with deep pressure in MRIPS #1. It appears as though a ceiling effect occurred for about half of the participants. As with coverage, data from MRIPS #3 was obtained from 7 of 8 participants. From MRIPS #2 to MRIPS #3, four of seven participants increased the area palpated at deep pressure, and three participants decreased the area palpated at deep pressure. Two of the participants whose performance decreased experienced large drop-offs, from 70% to 34% and from 54% to 31%. These participants may have experienced study fatigue. It is the opinion of our expert reviewers that these participants simply put less effort into palpation in MRIPS #3 than in the first two MRIPS exams.

Overall, from their first to last MRIPS interaction, 5 of 8 participants increased the area palpated at deep pressure and three participants decreased the area palpated at deep pressure. These three included the two participants who may have experienced study fatigue, and the participant who palpated the most area of any participants (95%, in MRIPS #1). Other than the two participants who put less effort into MRIPS #3 than MRIPS #2 and MRIPS #1, the repetitive practice with MRIPS appears to have a positive effect on learners’ performance in palpating with deep pressure. However, the ceiling effect and lack of effort on the part of two participants resulted in a lack of significant evidence of learning, and we must reject Hypothesis Pressure Learning.

9.6.3.4 Pressure transfer

From SP Pre-test to SP Post-test, participants significantly increased their use of the three levels of palpation pressure. In SP Pre-test, only 2 of 8 participants palpated with correct deep pressure, 3 of 8 palpated with correct medium pressure (but not deep
pressure), 2 of 8 palpated with correct light pressure (but not deep or medium), and one participant was rated as not palpating with even correct light pressure. In the SP Post-test, 6 of 8 participants palpated with correct deep pressure and 2 of 8 participants palpated with correct medium pressure. This represented an improvement for 5 participants and no change for 3 participants. These changes represented significant improvement, by a Wilcoxon test: $Z = 2.0, p = 0.03$. From this result, we accept Hypothesis Pressure Transfer. Appendix D.2 lists performance for each participant in coverage and pressure tasks.

### 9.6.4 Results: Impact of Real-Time Feedback on Coverage and Pressure

To evaluate whether the coverage and correct pressure feedback given by the touch map improved participants’ coverage and use of correct pressure in MRIPS, we compared participants’ coverage and correct pressure use in Study MRIPS-Learning with the previous Study MRIPSx2 (Section 4.3). This allows us to compare the performance of novices receiving real-time feedback to both novices and experienced residents and clinicians who did not receive real-time feedback.

The same mannequin, number of sensors, and density of sensors was used in both Study MRIPSx2 and Study MRIPS-Learning. The only significant difference with respect to palpation coverage and pressure was that a more precise method of evaluating coverage and pressure had been developed in the interim between the two studies. We re-processed the historical data to represent this more precise knowledge of what constitutes correct coverage and deep pressure.

The historical data from Study MRIPSx2 included unprocessed sensor value data from participants’ exams. In the previous evaluation in Section 4.3, coverage was evaluated as the area palpated at a pressure that was above the noise level for the
sensors. With the creation of the touch map feedback and evaluation (Section 7.4), we have a more precise definition for coverage: the percent area palpated at light or greater pressure. In order to generate coverage data that fit this more precise definition for coverage, we processed the historical data using a threshold which represented a lower bound for light coverage in Study MRIPS-Learning. This threshold was calculated as the minimum of the means for the light pressure distributions across the 64 sensors. Applying this threshold changed the percent area covered for only two of the Study MRIPSx2 participants, and by less than a significant amount (<2%).

A similar method was used to calculate the percent area palpated at deep or higher pressure in Study MRIPSx2. The threshold used for this calculation was the minimum of the means of the high pressure distributions across the 64 sensors. Because the method of calculating area in Study MRIPSx2 intentionally overestimates the area palpated, and the thresholds used in the calculations were lower bounds, the calculated areas of coverage and deep pressure use for the historical control should be considered upper bounds of actual performance by the historical control participants.

Because this between-subjects comparison had larger sample sizes, we used an independent-samples t-test to compare performance between Study MRIPSx2 and Study MRIPS-Learning. Results are shown in Table 9-8.

Participants in MRIPS #1 were performing their 2nd CBE ever, their first CBE being the SP Pre-test. Participants in the MRIPSx2 Inexperienced group were medical students with between 0 and 5 prior CBEs performed, meaning the MRIPS exam for which their coverage and pressure was evaluated was their 1st to 6th CBE. Participants in the MRIPSx2 Experienced group were experienced medical students, interns,
residents, and clinicians performing from their 7th to >1000th CBE (the experienced students had performed 6 to >10 CBES; most residents indicated they had performed between 50 and hundreds of CBES; clinicians indicated they had performed between hundreds and thousands of CBES).

Study MRIPS-Learning participants in MRIPS #1 performed significantly better at the coverage task, palpating significantly more breast tissue at light or higher pressure than either the MRIPSx2 inexperienced or MRIPSx2 experienced groups. The 12 participants in MRIPS #1 palpated 89.9% ± 11.6% of the breast at light or higher pressure, compared to 75.3% ± 12.4% for the MRIPSx2 Inexperienced group and 80.9% ± 9.0% for the MRIPSx2 Experienced group. The improvement over the MRIPSx2 inexperienced group is significant at $p < 0.005$ and the improvement over the MRIPSx2 experienced group is significant at $p < 0.05$. Both improvements are also significant at our ± 2% threshold.

Participants in Study MRIPS-Learning performing only their 2nd CBE were able to perform CBES with significantly more complete coverage than participants from Study MRIPSx2 with more prior experience in CBE. The independent variable that changed between Study MRIPS-Learning and Study MRIPSx2 is the presence of the touch-map real-time feedback of palpation completeness (coverage). From this result we accept Hypothesis Coverage Feedback Impact. The presence of the real-time touch-map feedback guides novice learners to expert-level (or better) performance in the cognitive-psychomotor task of palpating the entire breast.

The real-time touch-map feedback also resulted in improved performance in the psychomotor task of palpating with correct (deep) pressure. Participants in Study
MRIPS-Learning significantly outperformed the Study MRIPSx2 inexperienced participants. Study MRIPS-Learning participants palpated 73.7% ± 16.9% of the breast with deep pressure, compared to the 61.1% ± 17.5% percent palpated by the MRIPSx2 inexperienced group. This improvement was significant at α < 0.05. Study MRIPS-Learning participants also outperformed MRIPSx2 experienced participants, who averaged 62.9% ± 19.4% percent of tissue palpated at deep pressure. This improvement trended towards significance at p = 0.11. Though touch-map feedback did not lead Study MRIPS-Learning's novice participants to significantly outperform the more experienced historical control group, they did on average, and significantly outperformed the novice (approximately equally experienced) historical group. This result leads us to accept Hypothesis Pressure Feedback Impact. Receiving real-time feedback of the correctness of the learner’s palpation pressure causes learners to palpate more of the breast at correct deep pressure than when this feedback is not provided.

9.6.5 Results: Pattern-of-Search Learning and Transfer

In evaluating how closely participants followed the expert pattern in MRIPS, we analyzed only the most sensitive measure of deviation from the expert pattern, the normalized-deviation measure of Equation 9-2. Examples of patterns giving low and high normalized-deviation scores are shown in Figure 9-5. We expected to see the normalized-deviation decrease throughout the three MRIPS interactions. Performance is shown in Table 9-9. Seven of eight participants decreased their normalized-deviations from MRIPS #1 to MRIPS #2. This decrease is significant by a Wilcoxon test: Z = 2.2, p = 0.01. Of those improving, one participant switched from a spiral
pattern to the vertical strip pattern; the other participants improved the correctness of their vertical strip patterns.

From MRIPS #2 to MRIPS #3, total deviation increased for 5 participants and decreased for 2 participants. One participant did not have touch map or pattern-of-search map data for MRIPS #3. On average, there was an increase that trended towards significance, by Wilcoxon test: \( Z = 1.7, p = 0.06 \). However, the increase from MRIPS #2 to MRIPS #3 was much smaller than the decrease from MRIPS #1 to MRIPS #2.

The median change from MRIPS #1 to MRIPS #2 was a decrease of 5.7, while the median change from MRIPS #2 to MRIPS #3 was an increase of 1.9. The smaller increase from MRIPS #2 to MRIPS #3 following the larger decrease from MRIPS #1 to MRIPS #2 may indicate that some participants experienced a ceiling effect in MRIPS #2.

Overall, from participants’ first MRIPS interaction to their last MRIPS interaction, the total deviation decreased for 7 participants and increased for 1 participant. This decrease was significant at \( \alpha < 0.05 \), by a Wilcoxon test: \( Z = 1.96, p = 0.03 \).

Participants followed the expert’s pattern-of-search with significantly more precision after repeated practice with MRIPS incorporating pattern-of-search feedback, thus we accept Hypothesis Pattern-of-Search Learning.

Though the pattern-of-search feedback positively impacted learners’ performances, we did observe that some participants had problems with the pattern map related to errors in the tracking of the infrared-reflective marker on the participant’s middle fingernail. One problem was occlusion of the tracking marker. This most often
occurred when participants began palpating with their fingertips instead of the correct method of palpating with the finger pads. We observed two participants that experienced this problem. Participants also occluded the tracking marker by learning over the mannequin too far and blocking the camera’s view of the tracking marker. Participants were explicitly warned against this in the instructions before the interaction. Two other participants had this problem. These participants eventually figured out what was wrong, or asked the person supervising the study what was causing this problem. This issue did not impact the normalized or total deviation measures, as no segments were added to the pattern map while the marker was occluded. Participants were simply asked to start palpation over once the issue was fixed. The final problem was wearing of wristwatches or jewelry on the palpation hand. This occurred for two participants. This problem was more troubling as it caused errors in the pattern map. The watch face or jewelry reflected infrared light from the tracking camera which added noise to the calculated position of the tracking marker. Participants were instructed before the MRIPS interaction to remove any watches, jewelry, or reflective material from their hands, but these participants did not follow instructions. In our analysis of the normalized and total deviation measures, we removed segments which were obviously caused by this problem. These problems occurred for three participants in MRIPS #1, three participants in MRIPS #2, and for two participants in MRIPS #3. Solutions to these problems include having participants better follow instructions, increasing tracking infrastructure (e.g. a larger fiducial placed on the fingers that would not require the fingernail to face the camera), and adding more sophisticated tracking capabilities. Increasing tracking infrastructure is undesirable as it may negatively impact user
acceptability of MRIPS. However, tracking could be enhanced to track multiple distinct infrared blobs which could be used to separate the marker position from noise caused by other infrared-reflective objects in the scene.

To evaluate training transfer, we reviewed video of the SP interactions to evaluate what type of pattern-of-search participants used: vertical-strip pattern, other systematic pattern (e.g. spiral), or no systematic pattern.

In the SP Pre-test, two of eight participants used a vertical strip pattern, two used a spiral pattern, and four used a non-systematic (no discernable pattern) method of examining the breast. In the SP Post-test, seven of eight participants used a vertical strip pattern. One participant used a spiral pattern in SP Post-test. This participant also used the spiral pattern in SP Pre-test and MRIPS #1, then changed to a vertical strip pattern for MRIPS #2 and MRIPS #3, but reverted to the spiral pattern in the SP Post-test. The repetitive practice with the pattern-of-search feedback may not have been enough to ingrain in this participant that she should use a vertical strip pattern. However, the other seven participants did learn to use a vertical strip pattern. The change from two of eight (25%) using a vertical strip in SP Pre-test to seven of eight (~88%) using a vertical strip in SP Post-test is significant by McNemar’s test, $p = 0.03$.

Five of eight participants changed from a non-systematic or spiral pattern to a vertical strip pattern after repeated practice with MRIPS. All of these participants changed to a vertical strip pattern in MRIPS #1. The only information received by participants that instructed them to use a vertical strip pattern was the real-time pattern-of-search map feedback given in the MRIPS interactions. Given the significant improvement in the number of learners using vertical strip patterns and the evidence
that this change is due to the feedback provided in MRIPS, we accept Hypothesis Pattern-of-Search Transfer. Repeated practice guided by the real-time feedback present in MRIPS caused participants to use the correct vertical strip pattern instead of the incorrect patterns with which they began the study. Participants retained this knowledge learned in MRIPS and applied it to the CBE of the SP in the SP Post-test.

9.6.6 Results: Impact of Real-Time Feedback on Pattern-of-Search

To evaluate whether the presence of the pattern-of-search feedback guided participants to perform the vertical strip pattern-of-search, we compared the number of participants using the vertical strip pattern in MRIPS #1 to the number of participants using the vertical strip pattern in the historical control group of Study MRIPSx2 (Chapter 4). The 12 participants who completed MRIPS #1 were included in this analysis along with 18 inexperienced and experienced students in Study MRIPSx2. Eleven students from Study MRIPSx2 were not included in this analysis because video showing their patterns-of-search had not been recorded. From Study MRIPSx2, only students were included (not residents or clinicians) because the students had learned CBE technique at the Medical College of Georgia, where the vertical strip pattern is taught. Clinicians and residents learned CBE elsewhere and were taught a wider variety of techniques.

Ten of the twelve participants (83%) in MRIPS #1 used the vertical strip pattern, with two others using a systematic non-vertical strip pattern (horizontal strip and spiral patterns). From the historical control, 6 of 18 participants used a vertical strip pattern (33%), with 6 of 18 using a systematic non-vertical strip pattern (horizontal strip, spiral, and spoke patterns) and 6 of 18 using a non-systematic pattern (i.e. no discernable pattern). A significantly larger proportion of Study MRIPS-Learning participants used a vertical strip pattern than did the historical control group, by Fisher’s exact test at $p <$
The addition of the pattern-of-search map which was designed to guide learners to use a vertical strip pattern led to a significantly higher percentage of learners to use a vertical strip pattern. From this result, we accept Hypothesis Pattern-of-Search Feedback Impact.

We observed that two participants did not use a vertical strip pattern in MRIPS #1, despite receiving feedback that attempted to guide them to use a vertical strip pattern. Both of these participants used the same pattern they had previously used in the SP Pre-test. Only one of these participants completed MRIPS #2 and MRIPS #3; in these interactions, she used the vertical strip pattern. In MRIPS #1, neither participant attempted to follow the vertical strip pattern, so it appears unlikely that a high cognitive load made it too difficult for participants to follow the vertical strip pattern. As both participants used the same pattern as in the SP Pre-test, it seems plausible that they used these incorrect patterns out of familiarity – not from using these patterns in previous CBEs, but from misinformation from friends, attending physicians, or outdated educational materials. It is this type of non-standardized learning that MRIPS has the potential to combat, through repeated practice with standardized, objective feedback.

9.6.7 Results: Finding Masses and False Positives Learning and Transfer

Within the three MRIPS interactions, repeated practice with MRIPS appears to have positively impacted both finding of real masses and finding of false positive masses, but neither improvement was significant (Table 9-10). In MRIPS #1, two participants found one mass each, while in MRIPS #2 and MRIPS #3 the number of participants finding masses increased to three. In MRIPS #3, one of these three participants found both masses. This improvement was not significant. The number of participants and number of false positives found decreased with each repetition, but not
significantly. This was evaluated by a Friedman test, $X^2(df = 2, n = 8) = 2.0, ns$. From these tests, we must reject Hypothesis Finding Masses Learning and Hypothesis False Positive Masses Learning.

However, it is possible that with a larger population that these hypotheses would be accepted. This is because there is a relationship between finding masses and the amount of the breast palpated with deep pressure. Participants who found one or more real mass also palpated significantly more area with deep pressure than those participants who did not find a real mass. This difference trended towards significance in MRIPS #1 and was significant at $\alpha < 0.05$ in MRIPS #2 and MRIPS #3 (Table 9-11). This indicates that if MRIPS was able to improve learners’ palpation at deep pressure, then MRIPS should improve learners’ ability to find masses. We expect that with a larger population, and perhaps with more repetitions, a majority of participants would consistently palpate a high percentage (e.g. $>90\%$) of the breast with deep pressure. This would lead more participants to find the real masses. That we did not observe this significant improvement in finding masses is likely due to two participants performing very poorly in palpating with deep pressure ($<40\%$) in MRIPS #3. With these two participants not even attempting to palpate with enough pressure to find masses, only six participants remained who had the potential to find the masses. Fifty percent of participants attempting to find masses did find masses; if the same percentages held in a larger population, the increase from 25% (2 of 8) to 38% (3 of 8) to 50% (3 of 6) would have been significant.
Although there is insufficient evidence to accept that participants became more adept at finding masses due to repeated exposure to MRIPS, participants’ performance in finding masses and finding false positives improved from SP Pre-test to SP Post-test.

In the SP Pre-test, 4 of 8 participants found the real mass. All four also found the mass in the SP Post-test. Three additional participants did not find the mass in the SP Pre-test but did find the mass in the SP Post-test. The improvement from 50% of participants finding the mass to 87.5% of participants finding the mass trends towards significance by McNemar’s ($p = 0.13$). With a population ~1.5 times larger, the same proportions would be significant. It is worth noting then that if all 12 participants completing the SP Pre-test are included, the figure of 50% finding the mass in SP Pre-test still holds. However with only a trend towards significance, we will reject Hypothesis Finding Masses Transfer, but expect that a larger population would provide sufficient evidence to accept this hypothesis.

Though participants did not improve significantly in finding the real mass, participants improved significantly in distinguishing between normal tissue and masses, as evidenced by a significant reduction in false positives found in the SP breast. In the SP Pre-test, five false positives were found by five participants (mean $1.0 \pm 1.1$). This number decreased to three false positives found by two participants in SP Post-test (mean $0.38 \pm 0.74$). This decrease is significant by a Wilcoxon test at $\alpha < 0.05$: $Z = 2.2$, $p = 0.03$. From this evidence, we accept Hypothesis False Positive Masses Transfer.

Although we accepted only one of the four related hypotheses, practice with MRIPS does appear beneficial to the cognitive-psychomotor task of determining whether breast tissue is normal or a mass. The small population completing the study
did not provide the power to find a significant improvement in most of the tests, but incremental improvement was observed with repeated practice in MRIPS.

Overall, fewer participants found the masses in the MRIPS breast than in the SP breast. This indicates that the masses in the MRIPS breast were more difficult to palpate. This was expected of the soft mass, but expert clinicians rated the hard mass as being equally difficult to find as the SP mass. It appears that this was not the case for the novices who made up the study population. Although the majority of participants did not find a mass in MRIPS, we believe that the repeated practice with MRIPS resulted in the improvement from SP Pre-test to SP Post-test. Finding masses is a task in which performance does not improve without repetition and feedback [39]. These were provided by MRIPS with a frequency and precision otherwise unavailable in traditional curricula.

9.6.8 Discussion

Overall, participants’ performance in the psychomotor and cognitive-psychomotor components of the CBE improved with repeated practice with MRIPS. From SP Pre-test to SP Post-test, participants improved significantly in coverage, use of deep pressure, following the expert pattern-of-search, and not finding false positive masses. Participants also trended towards a significant improvement in finding real masses. Although the result of practice with MRIPS appears to be significant improvement in a CBE of an SP, ceiling effects and a limited population size made it difficult to evaluate the veracity of hypotheses of learning within MRIPS.

The real-time feedback appears to have contributed significantly in improving participants’ skills in these tasks, as participants in Study MRIPS-Learning significantly outperformed participants in Study MRIPSx2 (who did not receive real-time feedback).
Although the real-time feedback led to improved performance, it is likely the cause of the ceiling effects that prevented us from making more solid conclusions of learning.

Repeated practice with MRIPS appears to benefit learners’ performance in psychomotor and cognitive-psychomotor components of the CBE. However, learners appear to need more than three repetitions to learn some of the psychomotor tasks. After three opportunities to practice palpation at deep pressure, a large amount of variability in participants’ scores existed. Three participants palpated >90% and another at 88%, but the other four participants palpated <80% of the breast with deep pressure. This contrasts with the cognitive tasks in which the variance among participants’ scores decreased with each MRIPS interaction.

### 9.7 Affective Performance

Participants’ affective performance is primarily concerned with attending to the patient’s emotional state: displaying empathy and keeping the patient comfortable throughout the exam. In MRIPS, participants received feedback of their affective performances through the real-time thought-bubble feedback and post-experiential feedback which listed the number of successfully addressed, unsuccessfully addressed, and missed opportunities for comforting and expressing empathy. In the SP interactions, participants’ feedback concerning their affective performance came solely from verbal and nonverbal communication from the SP.

#### 9.7.1 Measures

In MRIPS, speech input was parsed to perform an objective evaluation of empathic content of participant responses to eight critical moments. We had intended for medical educators to provide video review of participant responses to these critical moments, but educators were unable to do so. Instead, participant handling of the
critical moments was rated objectively, assigning 0, 1, or 2 points to participant responses, based on a gross approximation of empathic content. In general, 2 points were assigned to responses which acknowledged the fear, concern, or emotional content of the patient’s statement; 1 point was assigned to responses which provided matter-of-fact related information; for all critical moments, 0 points were given to ignoring the participant or providing an unrelated response. The critical moments and specific scoring schemes are listed here:

- **CM1**: Before the physical exam, the patient asks in a fearful voice: “Do you think my pain could be because I have cancer?”
  - Scoring 2 points: acknowledging the patient’s concern, e.g. “from your mother passing away, I’m sure that must be a concern for you and I understand it must have been hard for you when you found the pain.”
  - Scoring 1 point: stating that a physical exam and tests are needed to determine if the pain is due to cancer, e.g. “it’s hard to say; are you willing to run some more tests to find out?”
- **CM2**: As the participant begins palpation, the patient expresses fear, that the exam will find cancer, in the form of a thought bubble: “This is so scary, what if they find cancer?”
  - Further interaction: If the participant responds by talking down to the patient, e.g. “there’s no need to be scared,” the participant responds with the thought: “Why shouldn’t I be scared? Of course I’m scared, what if it’s cancer?”
  - Further interaction: If the participant ignores the expression of fear, the patient responds with the thought “Maybe they only care about doing the exam, not about how I feel.”
  - Further interaction: If the participant expresses understanding, e.g. “I know this is hard for you because you lost your mother to breast cancer,” the participant responds with the thought: “I guess they do care about me, that makes me feel better.”
  - Scoring 2 points: acknowledging the patient’s fear or lack of comfort, e.g. “I understand this is uncomfortable.”
  - Scoring 1 point: explaining the exam, e.g. “I’m going to start by palpating along your collarbone.” No response received this score.
- **CM3**: After the physical exam is complete, the patient asks “Well, do you think it could be cancer?”
  - Scoring 2 points: acknowledging the patient’s concern, e.g. “I understand your concern about cancer with it running in your family. Soon we will be able to gather more information and tell you more certainly.”
Scoring 1 point: stating that more testing was required to determine if it could be cancer, e.g. “It’s hard to say that’s why we need to run a mammogram, to make sure it’s not cancer.”

CM4: When asked about her family history of cancer, the patient responds “I lost my mother to breast cancer two years ago. I miss her everyday.”

- Further interaction: If the participant ignores this opportunity for empathy, the patient thinks “I guess they just care about doing the exam.”
- Further interaction: If the participant’s response is on-topic but does not express understanding or sympathy, e.g. “How old was your mother when she died?” the patient thinks “This is my mother we’re talking about; this person doesn’t care at all.”
- Further interaction: If the participant expresses empathy (understanding, e.g. “I understand that must make you anxious about your breast pain. Have you been able to talk to anyone about how your mother’s passing is affecting you?”), or expresses sympathy (e.g. “I’m sorry to hear about your mother”), the patient responds with the thought “I guess they do care about me, that makes me feel better” and the speech “thank you, doctor.”
- Scoring 2 points: expressing empathy (understanding) or sympathy, e.g. “I know it’s hard to come in for something that your mother had as well and, not that you have breast cancer, but it’s a good decision to come see us” or “I am so sorry to hear that, it must be tough.”
- Scoring 1 point: asking more about her mother’s cancer, e.g. “how old was your mother when she passed away?”

CM5: When instructed that she should have a mammogram done, the patient expresses fear of mammograms: “I don’t know, I mean don’t mammograms hurt?”

- Scoring 2 points: acknowledging that the mammogram would be uncomfortable and/or reassuring the patient that she will be well taken care of, e.g. “I’ve heard they are uncomfortable but I actually don’t know if they are. The nurses there will do the best job they can to make it comfortable for you.”
- Scoring 1 point: stated that mammograms are important to diagnosis, e.g. “Well, we may need to do it just to make sure we rule out anything more serious.”

CM6: When further instructed that a mammogram is needed to ensure she does not have a malignancy, the patient expresses further fear of mammograms: “Do I really have to get a mammogram? I mean, my mom was fine, then she had a mammogram and all of the sudden she was really sick.”

- Scoring 2 points: acknowledging the patient’s fear, e.g. “I know it’s scary for you, but we need to do it so that we can figure out what’s going on with your breast pain.”
- Scoring 1 point: stating that the mammogram is important or referencing the loss of the patient’s mother as a motivating factor to get a mammogram, e.g. “mammograms are important diagnostic tools,” or “she probably had cancer before the mammogram; the mammogram simply found a cancer. It’s important to have a mammogram to detect it early so you can treat it.”
• CM7: When asked to disrobe for visual inspection, the patient expresses anxiety over disrobing, stating “I guess that’s ok… I’m kind of shy about taking off my clothes.”
  o Scoring 2 points: acknowledged the patient’s discomfort, e.g. “I understand. Please let me know at any time if you feel uncomfortable, OK?”
  o Scoring 1 point: stating why visual inspection should be performed, e.g. “right now I’m looking for any asymmetry or any redness or swelling.”
• CM8: When the patient was asked to raise her arms above her head for visual inspection, the patient thought: “This is so awkward; I don’t remember doing this for my last doctor.”
  o Scoring 2 points: acknowledging the patient’s discomfort. No response received this score.
  o Scoring 1 point: explaining why the patient needed to raise her arms over her head, e.g. “the reason I’m doing this is to make sure there is no mass in your armpit.”

Thought bubble feedback was also incorporated into virtual human responses to sexual history questions, but no participants queried sexual history, so we were not able to analyze that use of thought bubble feedback.

In evaluating learning, performance in the eight critical moments was condensed down to two measures. The first measure is the percent of moments scored a “2” (i.e. displaying empathy or acknowledging the patient’s emotions). The second measure is a normalized score of total performance, which is calculated by Equation 9-3, where \( n_x \) is the number of critical moments receiving a score of \( x \).

\[
Normscore = \frac{(2 \times n_2 + n_1)}{(n_2 + n_1 + n_0)}
\] (9-3)

For the SP interactions, affective performance was rated by medical experts and novices who reviewed video of participants’ exams. Affective performance was assessed by three measures using the instrument of Appendix D.4. These measures were an overall rating of empathy consisting of eight items scored on a 4-point forced-ranking scale (strongly disagree, disagree, agree, strongly agree). The other two measures concern the participant’s responses to four critical moments in which the
virtual human patient expressed concern or prompted for empathy. Participants were rated on the empathic content and appropriateness of their responses. The critical moments rated were:

- SP CM1: “I lost my mother to breast cancer two years ago. I miss her everyday.”
- SP CM2: “Do you think my pain could be because I have cancer.” (in a fearful voice, before physical exam)
- SP CM3: “Well, do you think it could be cancer?” (in a fearful voice, after physical exam)
- SP CM4: “Do I really have to get a mammogram? My mom was fine then she had a mammogram and then all of the sudden she was really sick?”

9.7.2 Hypotheses

- **Hypothesis Empathy Learning**: Participants’ demonstration of understanding and attentiveness to the patients’ emotions will improve with repeated MRIPS interactions.  
  - **Null hypothesis**: Affective performance scores will not significantly increase from MRIPS #1 to MRIPS #3.
- **Hypothesis Empathy Transfer**: From SP Pre-test to SP Post-test, participants will increase the appropriateness and empathic content of their responses to critical moments.  
  - **Null hypothesis**: Video reviewers’ ratings of the appropriateness and empathic content of participants’ critical moment responses will not significantly increase from SP Pre-test to SP Post-test.

9.7.3 Results: Empathy Learning

Performance is given for the three MRIPS-CBE interactions in Table 9-12, individual participant data is shown in Appendix D.6. From MRIPS #1 to MRIPS #2, four participants increased in the percent of moments rated a “2,” three participants decreased, and one participant did not change. This was not a significant change by Wilcoxon test: $Z = 0.74, p = 0.41$. From MRIPS #2 to MRIPS #3, four participants improved, one regressed, and three did not change their performance. This trended towards significant improvement by Wilcoxon test: $Z = 1.8, p = 0.063$. Overall, from MRIPS #1 to MRIPS #3, four participants improved, two regressed, and two did not
change; this was not a significant change in performance, by Wilcoxon test: $Z = 0.3, p = 0.4$.

In the normalized score, six participants improved from MRIPS #1 to MRIPS #2, one regressed and one did not change. This improvement trended towards significance, by Wilcoxon test: $Z = 1.4, p = 0.10$. From MRIPS #2 to MRIPS #3, six participants improved and two participants regressed. This improvement was not significant by Wilcoxon test: $Z = 0.84, p = 0.22$. Overall, from MRIPS #1 to MRIPS #3, seven participants improved and one did not change. This improvement was significant by Wilcoxon test at $\alpha < 0.01$: $Z = 2.4, p = 0.008$.

The normalized score is a more sensitive measure, though it also scores affective performance more liberally, i.e. some responses receiving a score of “1” were not empathetic – but were more appropriate than ignoring the patient in the critical moment. However, due to the difficulty in evaluating affective performance (earlier difficulties in getting medical education experts to agree in ratings of affective performance are described in Sections 4.2 and 4.3), and the small population size, we accept Hypothesis Empathy Learning on the basis of the significant improvement in the normalized score from MRIPS #1 to MRIPS #3.

9.7.4 Results: Impact of Feedback

Affective performance in MRIPS interactions prior to the incorporation of the thought bubble feedback was evaluated using different measures than affective performance in MRIPS in Study MRIPS-Learning. Thus we can not determine the impact of real-time feedback of affective performance by directly comparing performance with feedback and performance without feedback. Instead, we look at how participants reacted to the thought bubble feedback in Study MRIPS-Learning.
There were two critical moments (CM) in which virtual human communication consisted solely of thought bubbles: CM8 (awkwardness of visual inspection) and CM2 (fear that exam will find cancer). CM8 used a thought bubble to prompt a participant response. In CM8, only 3 of 7 participants (one did not do visual inspection and did not encounter CM8) reacted to the feedback (one participant in MRIPS #2 and two in MRIPS #3). The other four participants ignored the patient’s discomfort with visual inspection. The three who responded did so to explain the procedure, but did not acknowledge that the patient was uncomfortable with the procedure. In this instance, the thought bubble feedback appears ineffective at communicating the patient’s discomfort to the learner.

CM2 used a thought bubble to prompt participant response and a subsequent thought bubble to communicate to the participant whether his response (or lack of response) was appropriate. Only two participants responded to the prompt (one in MRIPS #1 and one in MRIPS #3). The one participant responding to the initial prompt in MRIPS #3 was the only participant who may have been motivated by negative thought bubble feedback (“maybe they only care about the exam”) to respond to the prompt in a subsequent interaction. With no participants responding to the prompt in two successive interactions, we can not determine whether positive feedback (“that makes me feel a bit better”) reinforced in the learner to continue to provide empathic responses.

Though we observed improvement in affective performance in MRIPS, we do not have any evidence that suggests the thought bubble feedback is directly responsible for improvement in affective performance. However, affective performance improved
significantly from MRIPS #1 to MRIPS #3 and from SP Pre-test to SP Post-test. This improvement may be due to post-experiential feedback (listing of successful and missed opportunities for empathy and video of expert reacting to three of the critical moments).

**9.7.5 Results: Empathy Transfer**

Participants improved their use of empathy significantly from SP Pre-test to SP Post-test. Results are shown in Table 9-12 with full data in Appendix D.5.

Participants were rated on the empathic content and appropriateness of their responses to the four critical moments listed in Section 9.7.1. We conducted a reliability analysis to determine if empathy and acceptability on each of the four critical moments could be averaged into unified empathy and acceptability scores that represented the overall empathy and acceptability of the participants’ responses to the critical moments. Cronbach’s alpha was calculated for empathy and acceptability measures in each SP interaction, and reliability for combining individual critical moment scores was high (α > 0.7): SP Pre-test empathy, α = 0.94; SP Pre-test appropriateness, α = 0.93; SP Post-test empathy, α = 0.98; SP Post-test appropriateness, α = 0.95. Thus, for each participant, the expert scores of the four critical moments were averaged. These scores are given in Table 9-12.

Participants improved significantly in their use of empathy when responding to critical moments. Only one of eight participants was rated positively on his use of empathy in response to critical moments in the SP Pre-test. All eight participants were rated positively on their use of empathy in response to critical moments in the SP Post-test. The improvement from a rating of 2.76 ± 0.65 to 3.95 ± 0.51 was significant at α < 0.005. The appropriateness of responses to critical moments also improved significantly, from 1 of 8 participants to 8 of 8 participants, with average rating improving
from 2.89 ± 0.56 to 4.02 ± 0.45 significant at α < 0.005. All participants improved in the empathic content and appropriateness of their handling of the patient’s concerns expressed in the critical moments.

Participants were also given an overall rating of empathy which took into account how the participant treated the patient’s emotions, the amount of concern the participant expressed for the patient, and the appropriateness of the participant’s non-verbal behavior. Participants improved significantly on this measure, with only 1 of 8 participants receiving a positive rating in the SP Pre-test and 6 of 8 participants receiving positive ratings in the SP Post-test. One of the two participants that did not receive an overall positive rating in the SP Post-test had a score of 2.9, just short of the minimum positive score of 3.0; this participant received positive scores from all reviewers except for one reviewer. The other participant not receiving a positive score in the SP Post-test had the same overall score in SP Pre-test and SP Post-test. This participant received overall positive scores in the critical moment measure, but was the only participant to receive any negative scores from individual reviewers in the critical moment ratings of the SP Post-test. It appears certain that 7 of the 8 participants improved their use of empathy and their handling of patient concerns as a result of repetitive practice with MRIPS. The overall empathy score improved significantly from SP Pre-test (2.58 ± 0.39) to SP Post-test (3.11 ± 0.38), at α < 0.05. From these results, we accept Hypothesis Empathy Transfer.

9.7.6 Discussion

Affective performance improved from repeated practice with MRIPS-CBE and this improvement transferred to the CBE of an SP. However, it is not clear what role the real-time and post-experiential feedback played in improving affective performance.
Due to the inability of medical educators to provide video review of performance in the MRIPS-CBE interactions, we were not able to directly compare affective performance to control groups. Further investigation should focus on evaluating the efficacy of the real-time feedback of affective performance.

9.8 Validity of Results

To evaluate the validity of the results of Study MRIPS-Learning, we investigated the impact of a single MRIPS-CBE interaction on performance in a CBE of an SP and the impact of an SP interaction on subsequent CBEs of SPs (a description of the evaluation was given in Section 9.2.3).

9.8.1 Impact of Multiple MRIPS Practice Opportunities

To determine the impact of a single MRIPS-CBE interaction, we compared performance of participants receiving a MRIPS-CBE practice opportunity before an evaluation in a CBE of an SP, and participants who did not receive an MRIPS-CBE practice opportunity before a CBE of an SP.

Cognitive performance was evaluated as the completeness of the breast history and completeness of visual inspection. Psychomotor performance was evaluated as completeness of palpation (coverage), use of correct palpation pressure, palpation of ancillary areas of tissue (supraclavicular, infraclavicular, and axilla), and use of a vertical strip pattern of search. Affective performance was evaluated using a four item instrument assessing empathy and appropriateness of verbal and nonverbal communication. Instruments and data are given in Appendix D.7. Results are shown in Table 9-13.

Participants that practiced with MRIPS-CBE performed significantly better on only the three psychomotor tasks. Real-time feedback was provided for two of these
tasks. The Study MRIPS-Learning participants performed better in the SP Post-test than the pilot study participants did in the SP interaction on all metrics other than coverage. However, the MRIPS-Learning participants only performed significantly better on the affective and cognitive tasks (Figure 9-14). That three MRIPS-CBE practice opportunities did not result in significantly better psychomotor performance than one MRIPS-CBE practice opportunity is likely due to the real-time feedback available for the coverage and palpation pressure tasks, as well as the small population size. Considering cognitive, psychomotor, and affective skill sets, there is clearly additional benefit to multiple practice opportunities incorporating real-time feedback of all three skill sets. However, this experiment does not tell us what number of repetitions will result in diminishing returns (when all participants’ performance plateaus).

9.8.2 Impact of an SP Pre-test Interaction on Subsequent SP Performance

To ensure that improvement from SP Pre-test to SP Post-test is not due solely to the practice afforded by the SP Pre-test, we compared performance in CBE of five medical students who performed two SP exams one month apart. For this group, the cognitive task of visual inspection, and the psychomotor tasks of coverage, palpating ancillary areas, using correct pressure, and using the vertical strip pattern-of-search were analyzed.

For visual inspection, four participants had the patient assume two or three visual inspection poses in their first SP interaction and decreased to doing no visual inspection in their second SP interaction; one participant used one pose in each SP interaction. The proportion of participants doing a visual inspection trended towards significant decrease from first to second SP interaction, by McNemar’s test: \( p = 0.063 \).
For coverage, two participants’ coverage decreased and three participants’ coverage stayed the same. The average coverage score decreased from 4.8 ± 1.1 to 3.6 ± 2.2, which was not significant by Wilcoxon test, Z = 1.3, p = 0.25. In the psychomotor task of palpating with correct pressure, two participants’ score decreased, one increased, and two remained the same. The average score decreased from 3.4 ± 1.3 to 3.0 ± 1.4, which was not significant, by Wilcoxon test, Z = 0.3, p = 0.5. Five of five participants palpated at least one ancillary area in their first SP interaction; this decreased to 2 of 5 participants in the second SP interaction. This decrease was not significant by McNemar’s test, p = 0.13. For the pattern-of-search, two participants did not use vertical strip in either SP interaction; two used vertical strip in the first SP interaction and did not use vertical strip in the second SP interaction; and one used vertical strip in both interaction. This decrease in performance was not significant by McNemar’s test.

All aspects of psychomotor performance decreased over the one month interval between first and second SP interactions. This is an approximately equal amount of time as between SP Pre-test and SP Post-test in Study MRIPS-Learning. If improvement in MRIPS-Learning from SP Pre-test to SP Post-test was due solely to the SP Pre-test experience, we would expect a similar improvement in this pilot study. Instead, we observed decreases in performance, indicating that improvement observed in Study MRIPS-Learning must be due to the three practice opportunities with MRIPS-CBE.

9.9 Study Limitations

The limitations of the study lie primarily in the small population. The population size was limited by the availability of medical student volunteers and SPs, as well as
Medical College of Georgia staff to administer the study. Many results trended towards significance and would likely be statistically significant in a larger population. Running a longer-duration study (~1 month from consent to completion) such as Study MRIPS-Learning – with a large population of medical students – encounters a catch-22: to obtain a large population, integration into curriculum is necessary; however, course integration requires that learning has previously been demonstrated, with such a study. Fortuitously, on the basis of Study MRIPS-Learning results, the Medical College of Georgia plans to incorporate multiple MRIPS-CBE interactions into their women’s health clerkship starting in April-May 2010. This will provide data for >100 students which will provide future insight into the efficacy of MRIPS for learning real-world interpersonal scenarios.

Additionally, it would be beneficial to compare learning with the current study design to learning in a sequence of five SP interactions. However, our medical collaborators express that practice with five SP interactions is unlikely in a medical curriculum and has prohibitive logistic and monetary costs. Although this would be an ideal control group for Study MRIPS-Learning, because we have previously validated MRIPS-CBE as a substitute for CBE of an SP, we should expect statistically similar impacts on performance from the Study MRIPS-Learning procedure and the ideal 5-SP procedure.

9.10 Revisiting Meta-Hypotheses

We restate the goals of Study MRIPS-Learning and corresponding meta-hypothesis and evaluate their veracity:

- Determine what learning (cognitive, psychomotor, affective) occurs in users of MRIPS-CBE.
- Meta-hypothesis: Participants will improve in cognitive, psychomotor, and affective performance throughout repetitive practice with MRIPS-CBE. Performance in these tasks will significantly improve from the first MRIPS interaction to the third MRIPS interaction.

- Accepted. Performance in psychomotor, cognitive, and affective tasks increased significantly with repeated practice in MRIPS-CBE. Specifically, performance improved significantly in breast history taking, visual inspection completeness, correctness of pattern-of-search, and affective tasks.

- Determine whether improvement in skills within MRIPS-CBE transfers to the real-world, in the form of improvement in performance in CBE of human patients.

- Meta-hypothesis: After practice with MRIPS-CBE, participants’ cognitive, psychomotor, and affective performance in CBE of an SP will have significantly improved in relation to baseline levels taken before practice with MRIPS-CBE.

- Accepted. Performance in psychomotor, cognitive, and affective tasks increased significantly from SP Pre-test to SP Post-test. Specifically, performance improved significantly on breast history completeness, visual inspection completeness, coverage, correctness of palpation pressure, correctness of pattern-of-search, not finding false positive breast masses, and appropriate use of empathy.

- Determine whether the presence of real-time feedback causes learners to significantly outperform past users of MRIPS-CBE (without real-time feedback) in cognitive, psychomotor, and affective tasks.

- Meta-hypothesis: Participants in Study MRIPS-Learning will perform significantly better in cognitive, psychomotor, and affective tasks (for which real-time feedback was provided) than prior users of MRIPS-CBE (participants in Study MRIPSx2, Section 4.3) who did not receive real-time feedback.

- Accepted for cognitive and psychomotor tasks. Real-time feedback resulted in improved performance in psychomotor, cognitive, and affective performance. Specifically, performance improved in the presence of real-time feedback for the tasks of breast history completeness, visual inspection completeness, coverage, correctness of palpation pressure, and correctness of pattern-of-search. The lack of a direct comparison to a control group left us unable to completely evaluate the impact of the thought bubble feedback. However, from within-subjects comparisons, the thought bubble feedback does not appear to be effective in improving affective performance. Yet, affective performance did improve; this may be due to the post-experiential affective feedback.
If learning occurs, determine whether skill sets are learned concurrently; or, whether one skill set must be maximized before performance in other skill sets can improve.

- Meta-hypothesis: We will not observe the following: performance in a single skill set requires maximization before the other skill sets can improve.

- Accepted. With the small population size and small number of repetitions, we provide an observational analysis. Trends in individual participants’ performance in the three skill sets are visualized in Table 9-16. Four participants clearly demonstrated concurrent learning of all three skill sets. Three participants demonstrated concurrent learning of two skill sets at a time. Two of these participants experienced no change in performance of the third skill set at all during the MRIPS interactions. This skill set was the cognitive skill set and the lack of change was due to ceiling effect in the breast history taking completeness. So, these participants also could be considered to learn all three skill sets concurrently; one skill set was simply maximized in the first practice opportunity. The final participant demonstrated concurrent learning of two skill sets, however this participant could be considered to have not learned concurrently, because he maximized affective performance before improving either cognitive or psychomotor performance. Overall, these observations indicate that MRIPS-CBE does afford concurrent learning of all three skill sets. This is made possible by the mix of haptics, patient interaction, and feedback that puts MRIPS at an advantage over other simulation approaches.

From Study MRIPS-Learning, we conclude that concurrent learning of cognitive, psychomotor, and affective skills takes place with repeated practice in MRIPS-CBE, and this improvement transfers to the real-world task of CBE of an SP. Real-time feedback plays a significant role in learning cognitive and psychomotor tasks of the breast exam, however, the role of feedback in affective learning is unclear. Though the study population was small, this study represents an initial step in demonstrating the efficacy of interpersonal simulation for educating real-world interpersonal scenarios.
Table 9-1. Instrument used to evaluate the completeness of breast history taking in MRIPS and SP interactions. Specific to the MRIPS interactions, the participant’s asking of shaded items was reviewed only in post-experiential feedback. Non-shaded items were visible during the CBE as part of the real-time procedural checklist feedback, and also reviewed in the post-experiential feedback.

<table>
<thead>
<tr>
<th>Breast history completeness instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief complaint</td>
</tr>
<tr>
<td>History of present illness: Location of pain</td>
</tr>
<tr>
<td>History of present illness: Description of pain</td>
</tr>
<tr>
<td>History of present illness: Nipple discharge</td>
</tr>
<tr>
<td>History of present illness: Other changes in breast (e.g. redness)</td>
</tr>
<tr>
<td>History of present illness: Trauma to breast</td>
</tr>
<tr>
<td>History of present illness: Pain stationary or radiating</td>
</tr>
<tr>
<td>History of present illness: Can the pain be made better (e.g. by medication)</td>
</tr>
<tr>
<td>Medical history: Age</td>
</tr>
<tr>
<td>Medical history: Onset of menarche (age of first menstrual period)</td>
</tr>
<tr>
<td>Medical history: Still having periods / are periods regular</td>
</tr>
<tr>
<td>Medical history: Use of hormones or birth control</td>
</tr>
<tr>
<td>Medical history: Prior breast problems</td>
</tr>
<tr>
<td>Medical history: Yearly clinical breast examination?</td>
</tr>
<tr>
<td>Medical history: Monthly self breast examination?</td>
</tr>
<tr>
<td>Medical history: Prior mammograms?</td>
</tr>
<tr>
<td>Medical history: Pregnancies</td>
</tr>
<tr>
<td>Medical history: Hospitalizations</td>
</tr>
<tr>
<td>Medical history: Surgeries</td>
</tr>
<tr>
<td>Medical history: Current medications</td>
</tr>
<tr>
<td>Medical history: Other health problems</td>
</tr>
<tr>
<td>Family History: History of cancer</td>
</tr>
<tr>
<td>Family History: Other medical problems</td>
</tr>
<tr>
<td>Social History: Smoker / tobacco use</td>
</tr>
<tr>
<td>Social History: Alcohol use</td>
</tr>
<tr>
<td>Social History: Sexually active?</td>
</tr>
<tr>
<td>Social History: Employment / on-the-job health risks</td>
</tr>
</tbody>
</table>

Table 9-2. Summary of results for hypotheses relating to cognitive performance.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast history</td>
<td>Learning and training transfer</td>
<td>Accepted</td>
</tr>
<tr>
<td>completeness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast history</td>
<td>Improvement from feedback</td>
<td>Accepted</td>
</tr>
<tr>
<td>completeness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual inspection</td>
<td>Learning and training transfer</td>
<td>Accepted</td>
</tr>
<tr>
<td>completeness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9-3. Performance in breast history taking in Study MRIPS-Learning.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Mean ± Std. dev.</th>
<th>Range of scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP Pre-test</td>
<td>12.0 ± 3.0</td>
<td>[8, 17]</td>
</tr>
<tr>
<td>MRIPS #1</td>
<td>20.3 ± 3.1</td>
<td>[15, 24]</td>
</tr>
<tr>
<td>MRIPS #2</td>
<td>20.1 ± 1.7</td>
<td>[18, 23]</td>
</tr>
<tr>
<td>MRIPS #3</td>
<td>21.3 ± 0.9</td>
<td>[20, 22]</td>
</tr>
<tr>
<td>SP Post-test</td>
<td>18.3 ± 3.9</td>
<td>[12, 25]</td>
</tr>
</tbody>
</table>

Table 9-4. Changes in the number of participants asking about specific risk factors. (*) Denotes that the increase is significant at $\alpha < 0.05$ by McNemar’s test, while (†) indicates that the increase trends to significance with $p = 0.06$.

<table>
<thead>
<tr>
<th>Item</th>
<th>SP Pre-test</th>
<th>SP Post-test</th>
<th>Importance [151]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset of menarche</td>
<td>0 of 8</td>
<td>5 of 8 *</td>
<td>Risk factor</td>
</tr>
<tr>
<td>Hormone / birth control use</td>
<td>1 of 8</td>
<td>7 of 8 *</td>
<td>Hormone use is a risk factor. Birth control can produce benign cysts which could cause the patient’s pain</td>
</tr>
<tr>
<td>Alcohol use</td>
<td>2 of 8</td>
<td>5 of 8</td>
<td>Risk factor with daily use</td>
</tr>
<tr>
<td>Job risks</td>
<td>1 of 8</td>
<td>5 of 8 †</td>
<td>Trauma may have caused the pain; exposure to environmental or chemical hazards increases risk</td>
</tr>
<tr>
<td>Prior breast problems</td>
<td>1 of 8</td>
<td>5 of 8 †</td>
<td>Risk factor</td>
</tr>
<tr>
<td>Patient age</td>
<td>3 of 8</td>
<td>7 of 8 †</td>
<td>Risk factor. Guides other questions e.g. concerning menopause and screening mammograms</td>
</tr>
</tbody>
</table>

Table 9-5. Number of participants performing any visual inspection and complete visual inspections in each interaction.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Performed Visual Inspection</th>
<th>Complete Visual Inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP Pre-test</td>
<td>2 of 8</td>
<td>0 of 8</td>
</tr>
<tr>
<td>MRIPS #1</td>
<td>6 of 8</td>
<td>5 of 8</td>
</tr>
<tr>
<td>MRIPS #2</td>
<td>7 of 8</td>
<td>4 of 8</td>
</tr>
<tr>
<td>MRIPS #3</td>
<td>7 of 8</td>
<td>5 of 8</td>
</tr>
<tr>
<td>SP Post-test</td>
<td>6 of 8</td>
<td>5 of 8</td>
</tr>
</tbody>
</table>
Table 9-6. Instrument used to evaluate coverage and use of correct pressure in the SP interactions. Coverage items were mutually exclusive (scores ranged from 0 to 6) and pressure items were cumulative (scores ranged from 0 to 4).

<table>
<thead>
<tr>
<th>Coverage items</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoroughly examined entire chest (6pts)</td>
<td></td>
</tr>
<tr>
<td>Thoroughly examined cone only (4 pts)</td>
<td></td>
</tr>
<tr>
<td>Did not examine cone thoroughly (0 pts)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pressure items</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used correct light pressure (1pt)</td>
<td></td>
</tr>
<tr>
<td>Used correct medium pressure (1pt)</td>
<td></td>
</tr>
<tr>
<td>Used correct deep pressure (2pts)</td>
<td></td>
</tr>
</tbody>
</table>

Table 9-7. Summary of acceptance and rejection of hypotheses of psychomotor and cognitive-psychomotor task performance.

<table>
<thead>
<tr>
<th>Task</th>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>Learning</td>
<td>For the entire breast, rejected (can not determine due to ceiling effect). Significant improvement in palpating supraclavicular, axilla, and infraclavicular areas.</td>
</tr>
<tr>
<td></td>
<td>Training transfer</td>
<td>Accepted</td>
</tr>
<tr>
<td></td>
<td>Improvement from feedback</td>
<td>Accepted</td>
</tr>
<tr>
<td>Pressure</td>
<td>Learning</td>
<td>Rejected (can not determine due to ceiling effect)</td>
</tr>
<tr>
<td></td>
<td>Training transfer</td>
<td>Accepted</td>
</tr>
<tr>
<td></td>
<td>Improvement from feedback</td>
<td>Accepted</td>
</tr>
<tr>
<td>Pattern-of-search</td>
<td>Learning</td>
<td>Accepted</td>
</tr>
<tr>
<td></td>
<td>Training transfer</td>
<td>Accepted</td>
</tr>
<tr>
<td></td>
<td>Improvement from feedback</td>
<td>Accepted</td>
</tr>
<tr>
<td>Finding masses</td>
<td>Learning</td>
<td>Rejected (improved but not significantly)</td>
</tr>
<tr>
<td></td>
<td>Training transfer</td>
<td>Rejected (trends towards significant improvement)</td>
</tr>
<tr>
<td>Reducing false positive masses</td>
<td>Learning</td>
<td>Rejected (improved but not significantly)</td>
</tr>
<tr>
<td></td>
<td>Training transfer</td>
<td>Accepted</td>
</tr>
</tbody>
</table>
Table 9-8. Coverage and use of deep pressure of the Study MRIPS-Learning participants in the MRIPS #1 interaction, compared to two historical control groups (from Study MRIPSx2 in Section 4.3) not receiving real-time feedback: inexperienced medical students (0 to 5 prior CBEs) and experienced medical students, residents, and clinicians (6 to >1000 prior CBEs). (*) Denotes results significant at $\alpha < 0.005$ and (**) denotes results significant at $\alpha < 0.005$.

<table>
<thead>
<tr>
<th></th>
<th>Coverage (percent area at light or higher pressure)</th>
<th>Pressure (percent area at deep or higher pressure)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± Stdev. (%)</td>
<td>95% CI</td>
</tr>
<tr>
<td>MRIPS #1 $(n = 12)$</td>
<td>89.9 ± 11.6</td>
<td>[82.6, 97.3]</td>
</tr>
<tr>
<td>MRIPSx2 Inexperienced $(n = 33)$</td>
<td>75.3 ± 12.4</td>
<td>[70.9, 79.6]</td>
</tr>
<tr>
<td>MRIPSx2 Experienced $(n = 24)$</td>
<td>80.9 ± 9.0</td>
<td>[77.1, 84.7]</td>
</tr>
<tr>
<td>MRIPS #1 vs. Inexperienced</td>
<td>$t(43) = 3.6$</td>
<td>$p = 0.001^{**}$</td>
</tr>
<tr>
<td>MRIPS #1 vs. Experienced</td>
<td>$t(34) = 2.6$</td>
<td>$p = 0.014^{*}$</td>
</tr>
</tbody>
</table>

Table 9-9. Total deviation from expert pattern-of-search in the three MRIPS interactions of Study MRIPS-Learning.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Mean ± Std. dev.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRIPS #1</td>
<td>22.6 ± 7.8</td>
<td>[17.2, 28.0]</td>
</tr>
<tr>
<td>MRIPS #2</td>
<td>16.8 ± 5.2</td>
<td>[13.2, 20.4]</td>
</tr>
<tr>
<td>MRIPS #3</td>
<td>19.3 ± 3.2</td>
<td>[17.1, 21.5]</td>
</tr>
</tbody>
</table>

Table 9-10. Number of participants finding real masses and false positive masses in MRIPS.

<table>
<thead>
<tr>
<th></th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masses found</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Participants</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>False positives</td>
<td>Participants</td>
<td>Masses</td>
<td>Participants</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 9-11. Participants finding masses palpated a larger percentage of the MRIPS breast with deep pressure. The difference is significant in MRIPS #2 and MRIPS #3.

<table>
<thead>
<tr>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>% area</td>
<td>n</td>
</tr>
<tr>
<td>Found 1+ masses</td>
<td>4</td>
<td>80.7 ± 17.0</td>
</tr>
<tr>
<td>Found no masses</td>
<td>8</td>
<td>70.2 ± 17.0</td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>Z = 1.2, p = 0.14</td>
<td>Z = 2.0, p = 0.029</td>
</tr>
</tbody>
</table>

Table 9-12. Affective performance in MRIPS-CBE interactions.

<table>
<thead>
<tr>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of moments scored “2”</td>
<td>23.4% ± 16.2%</td>
<td>20.8% ± 7.5%</td>
</tr>
<tr>
<td>Normalized score</td>
<td>0.59 ± 0.34</td>
<td>0.77 ± 0.26</td>
</tr>
</tbody>
</table>

Table 9-13. Expert ratings of participants’ affective performance in the SP interactions. (*) indicates significance at α < 0.05; (†) indicates significance at α < 0.005.

<table>
<thead>
<tr>
<th>Measure</th>
<th>SP Pre-test</th>
<th>SP Post-test</th>
<th>Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall empathy</td>
<td>2.58 ± 0.39</td>
<td>3.11 ± 0.38</td>
<td>Z = 2.24, p = 0.012†</td>
</tr>
<tr>
<td>Critical moment empathy</td>
<td>2.76 ± 0.65</td>
<td>3.95 ± 0.51</td>
<td>Z = 2.52, p = 0.004†</td>
</tr>
<tr>
<td>Critical moment appropriateness</td>
<td>2.89 ± 0.56</td>
<td>4.02 ± 0.45</td>
<td>Z = 2.52, p = 0.004†</td>
</tr>
</tbody>
</table>
Table 9-14. Performance in pilot study used to assess impact of a single MRIPS-CBE interaction on a subsequent CBE of an SP. Significant difference at $\alpha < 0.05$ is indicated by a (*) and trend towards significance is indicated by (**).

<table>
<thead>
<tr>
<th>Measure</th>
<th>MRIPS + SP ($n = 11$)</th>
<th>SP Only ($n = 8$)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast history</td>
<td>4.82 ± 1.2 (max score 10 items)</td>
<td>4.63 ± 2.0</td>
<td>Mann-Whitney: Z = 0.55, $p = 0.30$, ns</td>
</tr>
<tr>
<td>Performed any visual inspection</td>
<td>6 of 11 participants</td>
<td>5 of 8 participants</td>
<td>Fisher's exact test, ns</td>
</tr>
<tr>
<td>Coverage*</td>
<td>4.73 ± 1.0 (0 to 6 scale)</td>
<td>2.50 ± 2.1</td>
<td>Mann-Whitney, Z = 2.5, $p = 0.01$</td>
</tr>
<tr>
<td>Palpation pressure**</td>
<td>3.18 ± 1.4 (1 to 4 scale)</td>
<td>2.0 ± 1.8</td>
<td>Mann-Whitney, Z = 1.6, $p = 0.065$</td>
</tr>
<tr>
<td>Palpated one or more ancillary area*</td>
<td>8 of 11 participants</td>
<td>1 of 8 participants</td>
<td>Fisher's exact test, $p = 0.02$</td>
</tr>
<tr>
<td>Used vertical strip</td>
<td>6 of 11 participants</td>
<td>3 of 8 participants</td>
<td>Fisher's exact test, ns</td>
</tr>
<tr>
<td>Affective performance</td>
<td>3.06 ± 0.37 (max score 5.0; greater than 3.0 positive)</td>
<td>3.14 ± 0.50</td>
<td>Mann-Whitney, Z = 0.5, $p = 0.33$, ns</td>
</tr>
</tbody>
</table>

Table 9-15. Performance in a CBE of an SP after three MRIPS-CBE practice opportunities and after one MRIPS-CBE practice opportunity. Significant difference at $\alpha < 0.005$ is indicated by a (*)

<table>
<thead>
<tr>
<th>Measure</th>
<th>MRIPS-Learning participants after 3 MRIPS interactions (SP Post-test)</th>
<th>Pilot study participants after 1 MRIPS interaction</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast history*</td>
<td>7.86 ± 1.2 (max score 10 items)</td>
<td>4.82 ± 1.2</td>
<td>Mann-Whitney, Z = 3.5, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Visual inspection completeness (0-3)</td>
<td>2.0 ± 1.4</td>
<td>1.18 ± 1.33</td>
<td>Wilcoxon, Z = 1.3, $p = 0.14$, ns</td>
</tr>
<tr>
<td>Coverage</td>
<td>4.0 ± 2.6 (0 to 6 scale)</td>
<td>4.73 ± 1.0</td>
<td>Wilcoxon, Z = 0.09, $p= 0.49$, ns</td>
</tr>
<tr>
<td>Palpation pressure</td>
<td>3.75 ± 0.71 (1 to 4 scale)</td>
<td>3.18 ± 1.4</td>
<td>Wilcoxon, Z = 0.93, $p = 0.17$, ns</td>
</tr>
<tr>
<td>Ancillary completeness</td>
<td>1.5 ± 1.1</td>
<td>1.0 ± 0.78</td>
<td>Wilcoxon, Z = 1.2, $p = 0.13$, ns</td>
</tr>
<tr>
<td>Used vertical strip</td>
<td>7 of 8 participants</td>
<td>6 of 11 participants</td>
<td>Fisher's, $p = 0.15$, ns</td>
</tr>
<tr>
<td>Affective performance*</td>
<td>3.69 ± 0.44 (max score 5.0; positive greater than 3.0)</td>
<td>3.06 ± 0.37</td>
<td>Wilcoxon, Z = 2.7, $p = 0.002$</td>
</tr>
</tbody>
</table>
Table 9-16. Concurrent improvement in the three skill sets. For each MRIPS interaction, participants were rated as improving ("Up") in a skill set if they improved in >50% of the tasks of that skill set; improving in only 50% results in a rating of "NC" (no change). Participants 20, 33, 34, and 33 demonstrated concurrent learning of all three skill sets. Participants 21, 23, 27, and 37 demonstrated concurrent learning of two skill sets. An argument for lack of concurrent learning could be made for participant 37, as this participant maximized affective performance before improving cognitive and psychomotor performance.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>NC</td>
<td>NC</td>
<td>All three</td>
</tr>
<tr>
<td>21</td>
<td>Up</td>
<td>Dn</td>
<td>Up</td>
<td>NC</td>
<td>Up</td>
<td>Up</td>
<td>Two at a time</td>
</tr>
<tr>
<td>22</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>All three</td>
</tr>
<tr>
<td>23</td>
<td>NC</td>
<td>Dn</td>
<td>Dn</td>
<td>NC</td>
<td>Up</td>
<td>Up</td>
<td>Two at a time</td>
</tr>
<tr>
<td>24</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>Dn</td>
<td>Up</td>
<td>All three</td>
</tr>
<tr>
<td>27</td>
<td>NC</td>
<td>Dn</td>
<td>NC</td>
<td>NC</td>
<td>Up</td>
<td>Up</td>
<td>Two at a time</td>
</tr>
<tr>
<td>33</td>
<td>Up</td>
<td>Up</td>
<td>NC</td>
<td>Dn</td>
<td>Up</td>
<td>NC</td>
<td>All three</td>
</tr>
<tr>
<td>37</td>
<td>NC</td>
<td>Dn</td>
<td>Up</td>
<td>Up</td>
<td>Up</td>
<td>NC</td>
<td>No</td>
</tr>
</tbody>
</table>
Figure 9-1. Procedure for Study MRIPS-Learning.
Figure 9-2. Participants’ performance in breast history completeness in the two SP and three MRIPS interactions.
Figure 9-3. Real-time feedback appears to be more effective than the post-experiential feedback, as the gap in percentage-of-items-queried grew with repetition.
Figure 9-4. A) Participant drawing of correct locations of the two masses in the MRIPS breast. B) Participant drawing of correct location of the mass in the SP breast.

Figure 9-5. Participants’ use of deep pressure in MRIPS.
Figure 9-6. A) A pattern of search that closely followed the experts, receiving deviation scores of 72.1 (total) and 9.4 (normalized). B) A horizontal strip pattern did not closely follow the experts, receiving a total deviation score of 1270.6 and a normalized deviation score of 44.6. Note that the long non-outlined arrow is not part of the recorded pattern map nor is it included in the deviation score. This arrow points from the last palpation position to where the participant’s hand left the tracked area (at the end of the exam).
CHAPTER 10
SUMMARY AND FUTURE DIRECTIONS

We have developed a new paradigm for interpersonal simulation with virtual humans, in which the user’s touch of the virtual human and the user’s manipulation of physical objects are actively sensed by the virtual human and become significant components of the interaction between human and virtual human. Applying this paradigm to scenarios in medical education, we developed two mixed reality interpersonal simulations to train medical students’ cognitive, psychomotor, and affective skills in clinical breast examination and neurological examination with abnormal findings. We also demonstrated the importance of incorporating real-time feedback of user performance into interpersonal simulation.

10.1 Review of Results

In this dissertation we claimed that:

Interpersonal simulation incorporating instrumented haptic interfaces and providing real-time evaluation and feedback of performance improves users’ scenario-specific psychomotor, cognitive, and affective skills. Skills improvement transfers to the real-world interpersonal scenarios being simulated, demonstrated as improved performance in the real-world interpersonal scenario.

To evaluate this thesis statement, we first designed interfaces which allowed touch between human and virtual human and between virtual human and human, and allowed the human to interact with the virtual human through the manipulation of hand-held tools and hand-gestures.

We then evaluated the validity of the MRIPS approach for practicing and evaluating performance in two interpersonal scenarios in medical education. Content and construct validity were demonstrated for MRIPS-CBE, as cognitive, psychomotor, and affective performance in MRIPS-CBE were statistically equivalent or non-inferior to
performance with an SP; performance in MRIPS-CBE predicted performance with the
SP in cognitive and psychomotor tasks; and performance in MRIPS-CBE was able to
distinguish between inexperienced and experienced users in cognitive, psychomotor,
and affective tasks. Content validity was demonstrated for MRIPS-NEURO, as learners
were able to use MRIPS-NEURO to gather the information required to make a correct
diagnosis of the virtual human's cranial nerve disorder.

Real-time feedback was then developed to provide users of MRIPS with guidance,
reinforcement, and motivation for correction in cognitive, psychomotor, and affective
tasks. Real-time feedback in MRIPS-NEURO was shown to significantly improve
learner performance in the affective task of perspective taking and the efficiency of the
eye movements test, and the Patient-Vision feedback provided information sufficient for
learners to correctly diagnose the virtual human patient's cranial nerve disorder. Real-
time feedback in MRIPS-CBE was shown to significantly improve coverage, use of
correct pressure, use of a vertical strip pattern-of-search, and completeness of breast
history taking.

Learning and training transfer was evaluated in MRIPS-CBE. Overall, repeated
practice with MRIPS-CBE improved cognitive, psychomotor, and affective performance
in the CBE of standardized human patients. The small population and ceiling effects
limited the strength of these results, but it was clear that learning took place in MRIPS-
CBE and improvement in skills from practice with MRIPS-CBE transferred to the real-
world interpersonal scenario of CBE of a human patient.

10.2 Future Directions

The immediate next step for this work is integration of MRIPS-CBE and MRIPS-
NEURO into medical school curricula, so that more formal, larger studies of learning in
MRIPS can be conducted. Curriculum integration of MRIPS-CBE is planned to occur in the near future at the Medical College of Georgia and the University of Central Florida. MRIPS-NEURO is targeted for integration into the 1st-year medical school neuroscience course at the University of Florida. We also foresee the concept of haptic interaction with virtual humans being applied in two additional areas.

The first is a continuation of what we have begun with MRIPS. This is the integration of virtual humans with more advanced physical medical simulation, e.g. the Human Patient Simulator (HPS) [55]. By incorporating the medically-oriented inputs of the HPS (e.g. sensing of administered medication, chest compression, intubation, resuscitation) with the medially-oriented (palpation) and communication-oriented (speech, hand-gestures, tool use, comforting touches) inputs of MRIPS, and the expressiveness and flexibility of the virtual human, a myriad of medical procedures and scenarios could be simulated at a high degree of fidelity. The main challenge in the melding of virtual humans and the HPS is the display of the virtual human. While see-through head-mounted displays are an option, in our experience, HMDs present a barrier to widespread acceptability within the medical community. Pico-projectors or internal projection may provide a solution to the problem of displaying the virtual human.

The second is the application of haptic interaction with virtual humans beyond the medical domain, for example, training business greetings or a tangible virtual museum guide. Expansion into these domains will benefit from two future technological enhancements. One of these is a haptic interface that can move, e.g. walk, bow, shake hands, while remaining registered to the virtual human. The immediate solution is to incorporate more advanced robotics into a haptic interface similar to that of MRIPS-
The challenge lies in balancing the amount of robotics needed to perform the desired tasks with the amount of robotics visually acceptable to users. In other words, the addition of robotics and increased movement of the haptic interface should not cause the mixed reality human to fall into the uncanny valley. The other technological enhancement required to move into more general domains is the ability of the mixed reality human to interact with more than one human at a time. This will involve challenges in all aspects of design: enhancing speech recognition and understanding to recognize multiple speakers, tracking the position and attention (head gaze) of multiple users in order for the virtual human to address individual users, and the ability to distinguish among touches from multiple users. Within the development required to overcome these challenges, lies an opportunity for unique feedback mechanisms that expand on Patient-Vision feedback. Tracking multiple users will require the addition of more active sensors to the haptic interface, likely including stereo vision capabilities. Human-like stereo vision could be achieved by adding cameras at the two “eyes” of the virtual human and engineering the ability to have the cameras move along with the virtual human’s eyes. With this setup, Patient-Vision feedback of the real-world could be provided by manipulating the video streams of the two cameras and presenting the resulting images to one of the humans interacting with the virtual human. In this manner, the incorporation of additional sensing of user inputs (e.g. full body pose, proximity) and actuation of additional virtual human outputs (e.g. smell) will serve to further enhance the applicability of interpersonal simulation with virtual humans to a wide variety of real-world interpersonal scenarios.
APPENDIX A
STUDY MRIPS-SP DATA

A.1 Breast History Checklist Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>MRIPS (# items)</th>
<th>SP (# items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>104</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>107</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>202</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>203</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>204</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>205</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>206</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>207</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>

A.2 Empathy Video Review Instrument

1. Rate the appropriateness of the student's response -- how appropriate would this response be if the patient was real
   (1) Exceptionally INappropriate (one of the most inappropriate responses I can imagine)
   (2) Inappropriate
   (3) Borderline (could go either way)
   (4) Appropriate
   (5) Exceptionally appropriate (one of the most appropriate responses I can imagine)

Questions 2-4 are from the Empathy subscale of Krupat et al. [41]

2. Empathy: clinician allows patient to express emotions
   (1) Student shows no interest in patient's emotional state and/or discourages or cuts off the expression of emotion by the patient (verbal or nonverbal signals that it is not okay to express emotions).
   (2)
   (3) Student briefly acknowledges patient's feelings but makes no effort to indicate acceptance/validation.
   (4)
   (5) Student openly encourages or is receptive to the expression of emotion (e.g. through use of continuers or appropriate pauses (signals verbally or nonverbally that it is okay to express feelings).

3. Empathy: clinician validates patient feelings
   (1) Student makes no attempt to respond to/validate the patient's feelings, or possibly belittles or challenges them (e.g. It's ridiculous to be so concerned about...).
   (2)
   (3) Student briefly acknowledges patient's feelings but makes no effort to indicate acceptance/validation.
(4)
(5) Student makes comments clearly indicating acceptance/validation of patient's feelings (e.g. I'd feel the same way... I can see how that would worry you...).

4. Empathy: clinician explores patient feelings
   (1) Student makes no attempt to identify patient's feelings.
   (2)
   (3) Student makes brief reference to patients' feelings, but does little to explore them by identification or labeling.
   (4)
   (5) Student makes clear attempt to explore patient's feelings by identifying or labeling them (e.g. So how does that make you feel? It seems to me that you are feeling quite anxious about...)

5. Empathy: clinician nonverbal behavior
   (1) Student's nonverbal behavior displays lack of interest and/or concern and/or connection (e.g. little or no eye contact, body orientation or use of space inappropriate, bored voice).
   (2)
   (3) Student's nonverbal behavior shows neither great interest or disinterest (or behaviors over course of visit are inconsistent).
   (4)
   (5) Student displays nonverbal behaviors that express great interest, concern and connection (e.g. eye contact, tone of voice, and body orientation) throughout the visit.

A.3 Empathy Video Review Data

Critical moment: At the start of exam the patient exclaims “Wait! I'm scared” (7 experts rated this moment).

<table>
<thead>
<tr>
<th>ID</th>
<th>VP</th>
<th>SP</th>
<th>VP</th>
<th>SP</th>
<th>VP</th>
<th>SP</th>
<th>VP</th>
<th>SP</th>
<th>VP</th>
<th>SP</th>
<th>VP</th>
<th>SP</th>
<th>Avg. Emp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>208102</td>
<td>4.5</td>
<td>3.5</td>
<td>1.67</td>
<td>1.17</td>
<td>1.67</td>
<td>1.17</td>
<td>1.67</td>
<td>1.17</td>
<td>1.5</td>
<td>1.17</td>
<td>1.63</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>208103</td>
<td>2.17</td>
<td>3.84</td>
<td>1.33</td>
<td>1.5</td>
<td>1.17</td>
<td>1.33</td>
<td>1.17</td>
<td>1.33</td>
<td>1.5</td>
<td>1.21</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208104</td>
<td>4.0</td>
<td>4.0</td>
<td>1.33</td>
<td>1.5</td>
<td>1.17</td>
<td>1.33</td>
<td>1.17</td>
<td>1.33</td>
<td>1.5</td>
<td>1.5</td>
<td>1.33</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>208107</td>
<td>3.5</td>
<td>3.67</td>
<td>1.33</td>
<td>1.33</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.33</td>
<td>1.21</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208202</td>
<td>No data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208203</td>
<td>4.0</td>
<td>3.67</td>
<td>1.5</td>
<td>1.5</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.5</td>
<td>1.33</td>
<td>1.33</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>208204</td>
<td>4.17</td>
<td>3.33</td>
<td>1.5</td>
<td>1.5</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.5</td>
<td>1.33</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208205</td>
<td>4.17</td>
<td>3.0</td>
<td>1.33</td>
<td>1.33</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.33</td>
<td>1.33</td>
<td>1.25</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208206</td>
<td>3.67</td>
<td>3.33</td>
<td>1.5</td>
<td>1.5</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.5</td>
<td>1.33</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>208207</td>
<td>3.33</td>
<td>3.33</td>
<td>1.0</td>
<td>1.33</td>
<td>1.0</td>
<td>1.17</td>
<td>1.0</td>
<td>1.17</td>
<td>1.0</td>
<td>1.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg:</td>
<td>3.72</td>
<td>3.52</td>
<td>1.39</td>
<td>1.41</td>
<td>1.20</td>
<td>1.20</td>
<td>1.17</td>
<td>1.35</td>
<td>1.29</td>
<td>1.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kappa:</td>
<td>0.22</td>
<td>0.36</td>
<td>0.63</td>
<td>0.58</td>
<td>0.63</td>
<td>0.58</td>
<td>0.58</td>
<td>0.63</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Critical moment: Patient expresses a fear of cancer “Do you think it could be cancer?”
(6 experts rated this moment).

<table>
<thead>
<tr>
<th>ID</th>
<th>Appropriate</th>
<th>Empathy 1</th>
<th>Empathy 2</th>
<th>Empathy 3</th>
<th>Empathy 4</th>
<th>Avg. Emp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>208102</td>
<td>4.29</td>
<td>3.14</td>
<td>4.0</td>
<td>2.86</td>
<td>4.14</td>
<td>2.43</td>
</tr>
<tr>
<td>208103</td>
<td>3.43</td>
<td>4.0</td>
<td>3.14</td>
<td>2.43</td>
<td>3.14</td>
<td>2.57</td>
</tr>
<tr>
<td>208107</td>
<td>4.29</td>
<td>4.0</td>
<td>3.71</td>
<td>3.14</td>
<td>3.57</td>
<td>2.71</td>
</tr>
<tr>
<td>208202</td>
<td>2.71</td>
<td>3.0</td>
<td>1.86</td>
<td>2.57</td>
<td>1.71</td>
<td>2.14</td>
</tr>
<tr>
<td>208203</td>
<td>4.14</td>
<td>3.71</td>
<td>4.43</td>
<td>3.0</td>
<td>4.86</td>
<td>2.71</td>
</tr>
<tr>
<td>208205</td>
<td>3.57</td>
<td>3.71</td>
<td>3.0</td>
<td>2.71</td>
<td>3.0</td>
<td>2.29</td>
</tr>
<tr>
<td>208206</td>
<td>3.29</td>
<td>3.14</td>
<td>3.71</td>
<td>2.71</td>
<td>2.29</td>
<td>2.29</td>
</tr>
<tr>
<td>208207</td>
<td>2.57</td>
<td>3.43</td>
<td>1.71</td>
<td>2.29</td>
<td>1.71</td>
<td>2.43</td>
</tr>
<tr>
<td>Avg:</td>
<td>3.61</td>
<td>3.7</td>
<td>3.2</td>
<td>3.03</td>
<td>3.17</td>
<td>2.86</td>
</tr>
<tr>
<td>IRR:</td>
<td>0.30</td>
<td>0.43</td>
<td>0.15</td>
<td>0.24</td>
<td>0.31</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Critical moment: Explaining what will happen to patient “What happens next?”

<table>
<thead>
<tr>
<th>ID</th>
<th>Appropriate</th>
<th>Emp. 1</th>
<th>Emp. 2</th>
<th>Emp. 3</th>
<th>Emp. 4</th>
<th>Avg. Emp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>208102</td>
<td>4.5</td>
<td>4.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208103</td>
<td>2.73</td>
<td>3.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208104</td>
<td>4.33</td>
<td>4.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208105</td>
<td>4.17</td>
<td>3.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208202</td>
<td>3.33</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208203</td>
<td>2.67</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208204</td>
<td>4.5</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208205</td>
<td>3.5</td>
<td>3.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208206</td>
<td>3.67</td>
<td>2.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>208207</td>
<td>3.17</td>
<td>2.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg:</td>
<td>3.78</td>
<td>3.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRR:</td>
<td>0.18</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not applicable. This item was deemed by reviewers as not requiring empathy.
B.1 Study MRIPSx2 Video Review Instrument

1. Rate the appropriateness of the student's response -- how appropriate would this response be if the patient was real?
   (1) Exceptionally inappropriate (one of the most inappropriate responses I can imagine)
   (2) Inappropriate
   (3) Borderline (could go either way)
   (4) Appropriate
   (5) Exceptionally appropriate (one of the most appropriate responses I can imagine)

2. Rate the student's use of empathy...
   (1) Student was not at all empathetic
   (2) Student's attempt at empathy was not sincere (e.g. words are associated w/ empathy, but tone of voice demonstrates lack of sincerity)
   (3) Student's attempt at empathy was not successful (e.g. statement may not have been appropriate)
   (4) Student's empathy was appropriate and sincere
   (5) Student's empathy was exceptionally appropriate and sincere

B.2 Study MRIPSx2 Video Review Data

CM1: “Wait, I’m scared. What if you find cancer?”
CM2: “I lost my mother to breast cancer two years ago.”
CM3: “Do I really have to get a mammogram? I mean, my mom was fine, then she had a mammogram, and then all of the sudden she was really sick.”

Blank cells indicate that the participant did not experience this critical moment or that video reviewers did not evaluate this participant’s critical moment.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM1 Acceptable</td>
<td>3.4</td>
<td>4.1</td>
<td>3.4</td>
<td>4.1</td>
<td>4.5</td>
<td>4.7</td>
<td>3.5</td>
<td>3.7</td>
<td>4.2</td>
<td>3.3</td>
<td>2.7</td>
<td>4.0</td>
<td>4.0</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td>CM1 Empathetic</td>
<td>2.8</td>
<td>3.9</td>
<td>1.9</td>
<td>3.4</td>
<td>3.5</td>
<td>4.5</td>
<td>1.3</td>
<td>2.4</td>
<td>2.8</td>
<td>1.8</td>
<td>2.8</td>
<td>2.5</td>
<td>1.4</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>CM2 Acceptable</td>
<td>3.6</td>
<td>3.4</td>
<td>3.4</td>
<td>3.0</td>
<td>4.0</td>
<td>4.3</td>
<td>3.6</td>
<td>3.7</td>
<td>3.8</td>
<td>3.7</td>
<td>3.6</td>
<td>3.5</td>
<td>2.8</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>CM2 Empathetic</td>
<td>2.0</td>
<td>3.4</td>
<td>1.3</td>
<td>1.0</td>
<td>1.7</td>
<td>3.2</td>
<td>1.4</td>
<td>2.2</td>
<td>2.3</td>
<td>1.3</td>
<td>1.4</td>
<td>2.3</td>
<td>1.6</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>CM3 Acceptable</td>
<td>3.3</td>
<td>2.7</td>
<td>3.6</td>
<td>4.0</td>
<td>4.4</td>
<td>4.3</td>
<td>3.6</td>
<td>3.5</td>
<td>4.0</td>
<td>3.3</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>CM3 Empathetic</td>
<td>1.9</td>
<td>2.3</td>
<td>1.4</td>
<td>1.0</td>
<td>4.3</td>
<td>3.2</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>2.0</td>
<td>1.5</td>
<td>2.2</td>
<td>3.4</td>
<td>3.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>
B.3 Study MRIPSx2 Breast History Checklist Data

Blank cells indicate a clinician or resident (omitted from this analysis) or that a video or transcript was not available for this participant.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>701</th>
<th>702</th>
<th>703</th>
<th>704</th>
<th>707</th>
<th>708</th>
<th>709</th>
<th>713</th>
<th>714</th>
<th>716</th>
<th>722</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items asked</td>
<td>9</td>
<td>14</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>14</td>
<td>12</td>
<td>7</td>
<td>13</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>723</th>
<th>724</th>
<th>725</th>
<th>726</th>
<th>727</th>
<th>728</th>
<th>509</th>
<th>510</th>
<th>511</th>
<th>512</th>
<th>517</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items asked</td>
<td>8</td>
<td>5</td>
<td>12</td>
<td>11</td>
<td>6</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>518</th>
<th>519</th>
<th>521</th>
<th>523</th>
<th>524</th>
<th>525</th>
<th>526</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items asked</td>
<td>14</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

B.4 Study MRIPSx2 Palpation Completeness Data

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>701</th>
<th>702</th>
<th>703</th>
<th>704</th>
<th>707</th>
<th>708</th>
<th>709</th>
<th>710</th>
<th>712</th>
<th>713</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>94.1</td>
<td>89.5</td>
<td>72</td>
<td>70.9</td>
<td>96.5</td>
<td>85.9</td>
<td>89.4</td>
<td>70.9</td>
<td>75.5</td>
<td>94.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>714</th>
<th>716</th>
<th>718</th>
<th>722</th>
<th>723</th>
<th>724</th>
<th>725</th>
<th>726</th>
<th>727</th>
<th>728</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>47.1</td>
<td>87.1</td>
<td>90.1</td>
<td>76.2</td>
<td>83.6</td>
<td>74.5</td>
<td>82.5</td>
<td>92.9</td>
<td>73.3</td>
<td>83.7</td>
</tr>
</tbody>
</table>
APPENDIX C
STUDY MRIPS-NEURO QUESTIONNAIRES

C.1 Study MRIPS-NEURO Post-Patient Vision Survey

Post-Patient-Vision Survey

ID ______________________

1) Which eye was affected (or none)? __________________

2) Which cranial nerve was affected (or none)? _______________

3) Rate the patient’s ability to carry out everyday tasks
   (1 = needs assistance in all facets; 5 = normal ability):

   1  2  3  4  5

4) Briefly describe how you think the patient’s double vision affects his everyday life:

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

C.2 Study MRIPS-NEURO Post-Exam Survey

Post-Exam Survey

ID _____________________

1) For the Patient #1 (history and exam),
   Which cranial nerve was affected (or none)? _______________

   Which eye was affected (or none)? _________________

339
2) For the **Patient #2** (just exam),
Which cranial nerve was affected (or none)? ______________
Which eye was affected (or none)? _______________

3) List any concerns you have for **Patient #1** (history & exam), or anything you would like to express to the patient and/or his family:
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
## APPENDIX D

STUDY MRIPS-LEARNING INSTRUMENTS AND DATA

### D.1 Study MRIPS-Learning Breast History Checklist Data

<table>
<thead>
<tr>
<th>ID</th>
<th>SP Pre-test</th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
<th>SP Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>PE</td>
<td>Tot.</td>
<td>RT</td>
<td>PE</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>1</td>
<td>10</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>1</td>
<td>11</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>9</td>
<td>2</td>
<td>11</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>2</td>
<td>11</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>12</td>
<td>4</td>
<td>16</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>37</td>
<td>14</td>
<td>3</td>
<td>17</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>

Number of items queried by each participant. RT refers to items in the real-time procedural checklist. PE refers to items reviewed only in the post-experiential feedback. Tot. refers to the total number of items queried (Tot. = RT + PE).
### D.2 Study MRIPS-Learning Coverage and Pressure Data

<table>
<thead>
<tr>
<th>ID</th>
<th>Cvg</th>
<th>Press</th>
<th>% Cvg</th>
<th>%Deep</th>
<th>% Cvg</th>
<th>%Deep</th>
<th>% Cvg</th>
<th>%Deep</th>
<th>Press.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>90.1</td>
<td>42.4</td>
<td>81.9</td>
<td>51.2</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>1</td>
<td>99.1</td>
<td>82.1</td>
<td>95.7</td>
<td>80.7</td>
<td>99.2</td>
<td>87.8</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>2</td>
<td>72.1</td>
<td>65.5</td>
<td>89.2</td>
<td>82.0</td>
<td>99.2</td>
<td>91.1</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>1</td>
<td>98.0</td>
<td>94.7</td>
<td>94.2</td>
<td>84.2</td>
<td>99.6</td>
<td>88.5</td>
<td>4</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>2</td>
<td>66.0</td>
<td>62.1</td>
<td>91.2</td>
<td>70.0</td>
<td>86.6</td>
<td>34.2</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>4</td>
<td>95.0</td>
<td>62.7</td>
<td>86.4</td>
<td>51.3</td>
<td>89.2</td>
<td>74.1</td>
<td>4</td>
</tr>
<tr>
<td>33</td>
<td>6</td>
<td>4</td>
<td>76.8</td>
<td>66.4</td>
<td>98.3</td>
<td>91.6</td>
<td>90.3</td>
<td>76.7</td>
<td>6</td>
</tr>
<tr>
<td>37</td>
<td>0</td>
<td>0</td>
<td>92.8</td>
<td>76.0</td>
<td>89.9</td>
<td>53.5</td>
<td>91.1</td>
<td>31.2</td>
<td>4</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>2</td>
<td>99.5</td>
<td>96.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>2</td>
<td>96.1</td>
<td>58.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>0</td>
<td>2</td>
<td>98.8</td>
<td>94.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>2</td>
<td>95.0</td>
<td>83.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coverage and use of deep pressure in palpating the MRIPS and SP breast.

<table>
<thead>
<tr>
<th>ID</th>
<th>SP Pre-test</th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
<th>SP Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>37</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Palpation of supraclavicular, infraclavicular, and axilla. Number of the three areas palpated is indicated.
D.3 Study MRIPS-Learning Pattern-of-Search Data

<table>
<thead>
<tr>
<th>ID</th>
<th>SP Pre-test</th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
<th>SP Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Non-systematic</td>
<td>11.7</td>
<td>8.8</td>
<td>Missing data</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>21</td>
<td>Vertical strip</td>
<td>16.0</td>
<td>20.1</td>
<td>18.9</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>22</td>
<td>Spiral</td>
<td>16.4</td>
<td>9.4</td>
<td>14.1</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>23</td>
<td>Non-systematic</td>
<td>21.7</td>
<td>15.9</td>
<td>17.8</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>24</td>
<td>Non-systematic</td>
<td>35.3</td>
<td>18.9</td>
<td>21.1</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>27</td>
<td>Non-systematic</td>
<td>27.6</td>
<td>22.2</td>
<td>23.2</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>33</td>
<td>Vertical strip</td>
<td>23.6</td>
<td>17.9</td>
<td>22.4</td>
<td>Vertical strip</td>
</tr>
<tr>
<td>37</td>
<td>Spiral</td>
<td>28.4</td>
<td>21.4</td>
<td>21.3</td>
<td>Spiral</td>
</tr>
</tbody>
</table>

Normalized deviation of each participant’s pattern-of-search in MRIPS interactions and pattern used in SP interactions.

D.4 Study MRIPS-Learning Empathy Video Review Instrument

1. The participant did not pay attention to the patient’s emotions when interviewing and examining her (R).
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

2. The participant encouraged the patient to express her emotions.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

3. The participant accepted and/or validated the patient’s feelings.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

4. The participant displayed little interest or concern to the patient.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

5. The participant made little or no attempt to explore the patient’s feelings.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

6. The participant demonstrated appropriate non-verbal behavior.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
(4) Strongly agree
7. The participant dealt sensitively with embarrassing and disturbing topics and physical pain.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree
8. The participant expressed support and partnership.
   (1) Strongly disagree
   (2) Disagree
   (3) Agree
   (4) Strongly agree

The following two items were used to evaluate participant performance in four critical moments:
CM1: “I lost my mother to breast cancer two years ago.”
CM2: “Do you think my pain could be because I have cancer.” (before physical exam)
CM3: “Well, do you think it could be cancer?” (after physical exam)
CM4: “Do I really have to get a mammogram? My mom was fine then she had a mammogram and then all of the sudden she was really sick?”

1. The participant’s response was appropriate for a real patient.
   (1) Strongly inappropriate; offensive.
   (2) Less appropriate than average.
   (3) An average student might respond this way.
   (4) More appropriate than average.
   (5) Excellent; positive example to other students.
2. The participant’s response was appropriate for a real patient.
   (1) Strongly inappropriate; offensive.
   (2) Less appropriate than average.
   (3) An average student might respond this way.
   (4) More appropriate than average.
   (5) Excellent; positive example to other students.

D.5 Study MRIPS-Learning Empathy Video Review Data

<table>
<thead>
<tr>
<th>ID</th>
<th>Emp.1</th>
<th>Emp.2</th>
<th>Emp.3</th>
<th>Emp.4</th>
<th>Emp.5</th>
<th>Emp.6</th>
<th>Emp.7</th>
<th>Emp.8</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3.29</td>
<td>2.71</td>
<td>3.67</td>
<td>3.29</td>
<td>3.29</td>
<td>3.14</td>
<td>3.29</td>
<td>3.29</td>
<td>3.24</td>
</tr>
<tr>
<td>21</td>
<td>1.86</td>
<td>1.57</td>
<td>2.14</td>
<td>2.14</td>
<td>1.86</td>
<td>2.14</td>
<td>2.00</td>
<td>2.00</td>
<td>1.96</td>
</tr>
<tr>
<td>22</td>
<td>2.60</td>
<td>2.80</td>
<td>3.00</td>
<td>3.00</td>
<td>2.80</td>
<td>2.80</td>
<td>3.00</td>
<td>2.60</td>
<td>2.83</td>
</tr>
<tr>
<td>23</td>
<td>2.50</td>
<td>2.83</td>
<td>2.83</td>
<td>3.00</td>
<td>2.50</td>
<td>2.83</td>
<td>2.67</td>
<td>3.00</td>
<td>2.77</td>
</tr>
<tr>
<td>24</td>
<td>2.25</td>
<td>2.25</td>
<td>2.25</td>
<td>2.38</td>
<td>2.38</td>
<td>2.63</td>
<td>2.57</td>
<td>2.43</td>
<td>2.39</td>
</tr>
<tr>
<td>27</td>
<td>2.71</td>
<td>2.50</td>
<td>2.86</td>
<td>2.86</td>
<td>2.83</td>
<td>2.43</td>
<td>2.71</td>
<td>2.86</td>
<td>2.72</td>
</tr>
<tr>
<td>33</td>
<td>2.14</td>
<td>2.43</td>
<td>2.29</td>
<td>2.57</td>
<td>2.14</td>
<td>2.57</td>
<td>2.43</td>
<td>2.43</td>
<td>2.38</td>
</tr>
<tr>
<td>37</td>
<td>2.57</td>
<td>2.33</td>
<td>2.50</td>
<td>2.33</td>
<td>2.50</td>
<td>2.17</td>
<td>2.33</td>
<td>2.20</td>
<td>2.37</td>
</tr>
</tbody>
</table>
SP Pre-test data from the empathy instrument of Appendix D.4. Ratings greater than 3.0 are considered positive.

<table>
<thead>
<tr>
<th>ID</th>
<th>Emp.1</th>
<th>Emp.2</th>
<th>Emp.3</th>
<th>Emp.4</th>
<th>Emp.5</th>
<th>Emp.6</th>
<th>Emp.7</th>
<th>Emp.8</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3.57</td>
<td>3.00</td>
<td>3.33</td>
<td>3.50</td>
<td>3.17</td>
<td>3.33</td>
<td>3.00</td>
<td>3.17</td>
<td>3.26</td>
</tr>
<tr>
<td>21</td>
<td>3.50</td>
<td>3.00</td>
<td>3.38</td>
<td>3.75</td>
<td>3.25</td>
<td>3.38</td>
<td>3.38</td>
<td>3.5</td>
<td>3.39</td>
</tr>
<tr>
<td>22</td>
<td>4.00</td>
<td>3.29</td>
<td>3.57</td>
<td>3.71</td>
<td>3.29</td>
<td>3.29</td>
<td>3.57</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3.25</td>
<td>3.50</td>
<td>3.25</td>
<td>3.38</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.22</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2.75</td>
<td>3.00</td>
<td>3.25</td>
<td>3.00</td>
<td>3.13</td>
<td>3.00</td>
<td>3.13</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.00</td>
<td>2.67</td>
<td>3.17</td>
<td>3.14</td>
<td>2.86</td>
<td>2.71</td>
<td>2.83</td>
<td>2.91</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>2.50</td>
<td>2.29</td>
<td>2.43</td>
<td>2.29</td>
<td>2.14</td>
<td>2.14</td>
<td>2.43</td>
<td>2.31</td>
<td></td>
</tr>
</tbody>
</table>

SP Post-test data from the empathy instrument of Appendix D.4. Ratings greater than 3.0 are considered positive.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4.13</td>
<td>4.50</td>
<td>4.43</td>
<td>4.43</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.17</td>
<td>4.23</td>
</tr>
<tr>
<td>21</td>
<td>2.13</td>
<td>1.88</td>
<td>2.00</td>
<td>1.63</td>
<td>2.14</td>
<td>2.00</td>
<td>2.71</td>
<td>2.23</td>
<td>2.03</td>
</tr>
<tr>
<td>22</td>
<td>3.17</td>
<td>3.00</td>
<td>2.67</td>
<td>2.67</td>
<td>2.67</td>
<td>2.67</td>
<td>3.40</td>
<td>2.95</td>
<td>2.91</td>
</tr>
<tr>
<td>23</td>
<td>2.86</td>
<td>2.57</td>
<td>2.71</td>
<td>2.57</td>
<td>2.71</td>
<td>2.57</td>
<td>2.86</td>
<td>2.78</td>
<td>2.64</td>
</tr>
<tr>
<td>24</td>
<td>2.38</td>
<td>2.13</td>
<td>2.88</td>
<td>2.75</td>
<td>3.00</td>
<td>3.00</td>
<td>2.75</td>
<td>2.74</td>
<td>2.65</td>
</tr>
<tr>
<td>27</td>
<td>3.00</td>
<td>2.88</td>
<td>2.71</td>
<td>2.43</td>
<td>3.00</td>
<td>3.00</td>
<td>2.80</td>
<td>2.89</td>
<td>2.70</td>
</tr>
<tr>
<td>33</td>
<td>2.25</td>
<td>2.25</td>
<td>2.71</td>
<td>2.43</td>
<td>2.71</td>
<td>2.43</td>
<td>2.50</td>
<td>2.53</td>
<td>2.40</td>
</tr>
<tr>
<td>37</td>
<td>3.25</td>
<td>2.75</td>
<td>2.86</td>
<td>2.43</td>
<td>2.71</td>
<td>2.57</td>
<td>2.20</td>
<td>2.81</td>
<td>2.52</td>
</tr>
</tbody>
</table>

SP Pre-test critical moment appropriateness and empathy ratings.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4.57</td>
<td>4.00</td>
<td>4.40</td>
<td>4.40</td>
<td>4.50</td>
<td>4.50</td>
<td>4.40</td>
<td>4.47</td>
<td>4.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>4.50</td>
<td>4.50</td>
<td>4.67</td>
<td>4.67</td>
<td>4.71</td>
<td>4.71</td>
<td>4.60</td>
<td>4.80</td>
<td>4.63</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3.43</td>
<td>3.57</td>
<td>4.00</td>
<td>3.80</td>
<td>4.00</td>
<td>3.80</td>
<td>4.14</td>
<td>4.00</td>
<td>3.88</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3.75</td>
<td>3.88</td>
<td>3.86</td>
<td>3.86</td>
<td>3.88</td>
<td>3.88</td>
<td>4.00</td>
<td>3.88</td>
<td>3.87</td>
<td>3.87</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>4.00</td>
<td>4.00</td>
<td>4.14</td>
<td>4.29</td>
<td>4.00</td>
<td>4.25</td>
<td>3.75</td>
<td>3.75</td>
<td>3.97</td>
<td>4.06</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.25</td>
<td>3.50</td>
<td>3.71</td>
<td>3.57</td>
<td>3.50</td>
<td>3.38</td>
<td>4.14</td>
<td>3.71</td>
<td>3.63</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>3.14</td>
<td>3.00</td>
<td>3.20</td>
<td>3.00</td>
<td>3.25</td>
<td>3.00</td>
<td>3.75</td>
<td>3.25</td>
<td>3.30</td>
<td>3.05</td>
<td></td>
</tr>
</tbody>
</table>

SP Post-test critical moment appropriateness and empathy ratings.
D.6 Affective Ratings of Participants in MRIPS-CBE

<table>
<thead>
<tr>
<th>ID</th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
<th>MRIPS #1</th>
<th>MRIPS #2</th>
<th>MRIPS #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>14.3</td>
<td>25.0</td>
<td>25.0</td>
<td>0.57</td>
<td>1.00</td>
<td>0.63</td>
</tr>
<tr>
<td>21</td>
<td>33.3</td>
<td>28.6</td>
<td>37.5</td>
<td>0.67</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>22</td>
<td>25.0</td>
<td>28.6</td>
<td>25.0</td>
<td>0.75</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>23</td>
<td>37.5</td>
<td>12.5</td>
<td>25.0</td>
<td>1.00</td>
<td>0.38</td>
<td>1.00</td>
</tr>
<tr>
<td>24</td>
<td>0.00</td>
<td>14.3</td>
<td>33.3</td>
<td>0.00</td>
<td>0.43</td>
<td>0.83</td>
</tr>
<tr>
<td>27</td>
<td>50.0</td>
<td>28.6</td>
<td>33.3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.17</td>
</tr>
<tr>
<td>33</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>0.50</td>
<td>0.63</td>
<td>0.88</td>
</tr>
<tr>
<td>37</td>
<td>14.3</td>
<td>16.7</td>
<td>16.7</td>
<td>0.29</td>
<td>1.00</td>
<td>0.50</td>
</tr>
</tbody>
</table>

D.7 Pilot Study Video Rating Instrument and Data

   a. Pain
   b. Discharge
   c. Mass
   d. Breast self examination
   e. Family history
   f. Alcohol use
   g. Smoking history
   h. Birth control pills / Medications
   i. Age of first pregnancy
   j. Onset of period

2. Affective
   a. Attentiveness (interest in patient’s problems)
      (1) Did not seem to really be paying attention or listening; interrupted without apology of explanation.
      (2) Attention drifted at times; asked a question that had already been answered without apology.
      (3) Appeared to be paying attention.
      (4) Appeared to be paying attention and responded to verbal or non-verbal cues.
   b. Eye contact (appropriateness of nonverbal behavior)
      (1) Little or no eye contact
      (2) Some eye contact
      (3) Appropriate eye contact at most times
      (4) Appropriate eye contact at all times
   c. Attitude (appropriateness of verbal behavior)
      (1) Made judgmental comments or criticized patient; or talked down to patient.
      (2) Made 1-2 comments with inappropriate affect.
(3) No judgmental comments; talked to patient as an equal.
(4) No judgmental comments; talked to patient as an equal; offered praise/encouragements when opportunity arose.

d. Empathy and Support
   (1) Offered no empathetic comments; no encouragement or support (did not state intention to help).
   (2) Offered only brief supportive or empathetic comment and only in response to a distinct emotional statement by patient. Comments may seem prospective or forced.
   (3) Offered empathetic or supportive comments or stated intention to help.
   (4) Offered empathetic or supportive comments or stated intention to help; despite limited time seemed to be on the way to establishing a caring relationship.

3. Psychomotor: physical exam.
   a. Inspected both breasts arms relaxed (0 or 1).
   b. Inspected both breasts arms flexed (0 or 1).
   c. Inspected both breasts arms raised (0 or 1).
   d. Examined supraclavicular area (0 or 1).
   e. Examined infraclavicular area (0 or 1).
   f. Examined deep central axillary area (0 or 1).
   g. Use of pressure
      • Light only: 1 pt.
      • Light and medium: 2 pts.
      • Deep (and medium and light): 4 pts.
   h. Used vertical strip pattern (0 or 1).
   i. Coverage
      • Examined entire chest area: 6 pts.
      • Examined cone only: 4 pts.
      • Incomplete examination of cone: 0 pts.
<table>
<thead>
<tr>
<th>ID</th>
<th>Visual Inspection (0 to 3)</th>
<th>Breast History (0 to 10)</th>
<th>Ancillary Coverage (0 to 3)</th>
<th>Coverage (0 to 6)</th>
<th>Pressure (0 to 4)</th>
<th>Vertical Strip</th>
<th>Affective (1 to 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>No</td>
<td>2.38</td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>Yes</td>
<td>3.25</td>
</tr>
<tr>
<td>51</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>2.83</td>
</tr>
<tr>
<td>53</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>No</td>
<td>3.63</td>
</tr>
<tr>
<td>49</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>Yes</td>
<td>3.25</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>No</td>
<td>2.53</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>Yes</td>
<td>3.68</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>Yes</td>
<td>3.56</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>Yes</td>
<td>3.10</td>
</tr>
<tr>
<td>91</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>No</td>
<td>2.90</td>
</tr>
<tr>
<td>73</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>No</td>
<td>3.35</td>
</tr>
<tr>
<td>64</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>Yes</td>
<td>2.40</td>
</tr>
<tr>
<td>63</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>No</td>
<td>3.38</td>
</tr>
<tr>
<td>69</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>No</td>
<td>3.45</td>
</tr>
<tr>
<td>68</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>No</td>
<td>2.75</td>
</tr>
<tr>
<td>74</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>3.00</td>
</tr>
<tr>
<td>54</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>Yes</td>
<td>3.25</td>
</tr>
<tr>
<td>70</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>Yes</td>
<td>3.50</td>
</tr>
<tr>
<td>83</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>2.57</td>
</tr>
<tr>
<td>56-2nd SP</td>
<td>0</td>
<td>n/a</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>51-2nd SP</td>
<td>0</td>
<td>n/a</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>Yes</td>
<td>n/a</td>
</tr>
<tr>
<td>47-2nd SP</td>
<td>1</td>
<td>n/a</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>46-2nd SP</td>
<td>0</td>
<td>n/a</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>73-2nd SP</td>
<td>0</td>
<td>n/a</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>No</td>
<td>n/a</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Aaron Andrew Kotranza was born in 1983 in Tampa, Florida. Aaron graduated cum laude from Berkeley Preparatory School in 2001. He received a full scholarship to attend the University of Florida as a National Merit Scholar. Aaron began working with Dr. Benjamin Lok as an undergraduate in 2004. Upon graduating with honors from the undergraduate computer engineering program, Aaron was awarded a four-year UF Alumni Fellowship to pursue a Ph.D. His work has focused on expanding the applicability of virtual humans to simulate social scenarios for interpersonal skills training. His work has received significant recognition in both the fields of computer science and medicine with 11 articles published in leading journals and conferences, including receiving the best paper award at IEEE Virtual Reality 2008 and a featured article in the May/June 2009 issue of *IEEE Transactions on Visualization and Computer Graphics*. Aaron and his collaborators have applied for patents of the technology described in this dissertation. In the near future, Aaron hopes to continue his work in a research position in academia or industry.