
Communicative Humanoids

A Computational Model of Psychosocial Dialogue Skills

Kristinn Rúnar Thórisson

B.A. *Cognitive Psychology*, University of Iceland, 1987

M.S. *Engineering Psychology*, Florida Institute of Technology, 1990

*Submitted to the Program in Media Arts & Sciences,
School of Architecture & Planning,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the **Massachusetts Institute of Technology***

September 1996

© 1996, Massachusetts Institute of Technology

Author:

Program in Media Arts & Sciences

July 19, 1996

Certified by:

Justine Cassell

Assistant Professor of Media Arts & Sciences

MIT Program in Media Arts & Sciences

Thesis Advisor

Accepted by:

Stephen A. Benton

Chair, Departmental Committee on Graduate Students

Program in Media Arts & Sciences

Communicative Humanoids

A Computational Model of Psychosocial Dialogue Skills

Kristinn Rúnar Thórisson

*Submitted to the Program in Media Arts & Sciences,
School of Architecture & Planning on July 19, 1996
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at the Massachusetts Institute of Technology*

Abstract

Face-to-face interaction between people is generally effortless and effective. We exchange glances, take turns speaking and make facial and manual gestures to achieve the goals of the dialogue. Endowing computers with such an interaction style marks the beginning of a new era in our relationship with machines—one that relies on communication, social convention and dialogue skills. This thesis presents a computational model of psychosocial dialogue expertise, bridging between perceptual analysis of multimodal events and multimodal action generation, supporting the creation of interfaces that afford full-duplex, real-time face-to-face interaction between a human and autonomous computer characters. The architecture, called *Ymir*, has been implemented in software, and a prototype humanoid created. The humanoid, named *Gandalf*, commands a graphical model of the solar system, and can interact with people using speech, manual and facial gesture. *Gandalf* has been tested in interaction with users and has been shown capable of fluid face-to-face dialogue. The prototype demonstrates several new ideas in the creation of communicative computer agents, including *perceptual integration of multimodal events*, *distributed processing* and *decision making*, *layered input analysis* and *motor control*, and the integration of *reactive* and *reflective perception* and *action*. Applications of the work presented in this thesis can be expected in such diverse fields as education, psychological and social research, work environments, and entertainment.

Thesis Advisor:

Justine Cassell

Assistant Professor of Media Arts & Sciences
MIT Program in Media Arts & Sciences

This research was sponsored by the Media Laboratory and Thomson-CSF.

Doctoral Dissertation Committee

Thesis Advisor:

Justine Cassell

Assistant Professor of Media Arts & Sciences

M.I.T. Program in Media Arts & Sciences

Thesis Reader:

Pattie Maes

Associate Professor of Media Arts & Sciences

Sony Corporation Career Development Professor
of Media Arts & Sciences

M.I.T. Program in Media Arts & Sciences

Thesis Reader:

Steve Whittaker

Research Scientist

AT&T Bell Laboratories

Acknowledgments

A number of people helped shape and polish the ideas expressed in this thesis and they deserve credit. I want to thank Justine Cassell, for her unique perspective, professional advice, and for continued support through the completion of this thesis. I thank my readers for their recommendations, expertise and moral support; Pattie Maes, for her initial interest in this work, continued encouragement, and for feeding me with material that proved essential for its completion; Steve Whittaker for pointing out all those important details and for his keen observations from the psychological perspective. Thomas Malone, who was a member on my general exam committee, brought to my attention important issues in system architecture and design. Richard A. Bolt gave invaluable support in the beginning stages.

Thanks to the members of the Gesture & Narrative Language Group, my “new” gang: Jennifer Glos, Marina Umachi, Hannes Vilhjálmsson, Scott Prevost and Erin Panttaja, whose comments, suggestions and work has helped me shape this thesis, and who, by being there, made finishing up infinitely more fun.

Thanks to the students of the past who made life considerably more fun here at the lab, and some of whom I had the additional pleasure of working with on a project or two: Dave Berger, Brent Britton, Karen Donoghue, Tony Ezzat, Ed Herranz, Jen Jacoby, Matt Kaminsky, Alice Lei, Steve Librande, Mark Lucente, Bob Sabiston, David Small, Carl Sparrell, and Chris Wren.

My undergraduate research assistants, who collectively detailed the inner workings of ToonFace: Steven Levis, Roland Paul, Nimrod Warschawsky and Calvin Yuen.

Thanks to my office mate, Joshua Bers, who is always ready to discuss anything from the general to the particular, but most importantly my dedicated partner in all of the most wacky late-night madness generated during long hours of debugging. And to Tomoko Koda for livening up the place.

Thanks to the volunteers in my human subjects study.

Thanks to Bebbá and Jóel for donating me the Sanyo Z1 cassette/CD player boom box. Couldn't have made without it!

Thanks to the cheerful crowd on the second floor, especially Linda Peterson, for help and understanding. To the syspro guys for always being on call.

Thanks to Nicholas for the cappuchino machine next to my office.

Thanks to my father, fiórir S. Guþbergsson, and mother, Rúna Gísladóttir, for invaluable support and for providing me with the skills and will to enjoy my work.

But most of all, thanks to my wife Katrín Elvarsdóttir, whose dedication made it possible for me to do this work.

Time to do the laundry.

To
Katrín Elvarsdóttir

Abstract 3

Acknowledgments 7

1.

Introduction 19

- 1.1 What is Needed 20
- 1.2 Goals of This Work 21
 - Terms & Definitions* 22
 - Outline of Thesis* 23

2.

Face-to-Face Interface 25

- 2.1 Humanoid Agents: Early History 25
- 2.2 Face-to-Face: When & Why 26
 - Some Compelling Reasons for Interacting Face-to-Face* 28
 - Face-to-Face: When NOT?* 32
 - Anthropomorphization: A Non-Traditional Perspective* 34
- 2.3 Summary 35

3.

Multimodal Dialogue: Psychological and Interface Research 37

- 3.1 Human Multimodal Communication 38
 - Dialogue Structure* 38
 - Turn Taking* 38
 - Back-Channel Feedback* 40
 - Embodied Conversants* 41
 - The Multiple Modes of Face-to-Face Interaction* 42

- 3.2 Multimodal Computer Interfaces 47
 - Multimodal Analysis and Interpretation* 49
 - Missing Pieces in the Multimodal Metaphor* 51

4.

Agents, Robots & Artificial Intelligence 53

- 4.1 The Agent Metaphor 53
 - Agent Embodiment* 55
 - Visual Representation* 56
 - Spatial Representation* 57
- 4.2 Agent Architectures 59
 - Classical A.I.* 60
 - Behavior-Based A.I.* 61
 - Hybrid Systems* 62
- 4.3 Summary 63

5.

Computational Characteristics of Psychosocial Dialogue Skills 65

- 5.1 Challenges of Real-Time Multimodal Dialogue 66
- 5.2 Temporal Constraints 69
- 5.3 Functional Analysis: A Precursor to Content Interpretation and (sometimes) Feedback Generation 71
 - The Link Between Functional Analysis and Process Control* 73
- 5.4 Turn Taking 74
 - A Situated Model of Turn Taking* 75
- 5.5 Morphological and Functional Substitutability 76

- 5.6 Multimodal Dialogue as Layered Feedback
Loops 77
- 5.7 Summary 79

6. *J.Jr.: A Study in Reactivity* 81

- 6.1 System Description 81
 - Input: Gestures, Gaze & Intonation* 81
 - Output: Speech, Turn Taking, Back Channel, Gaze* 82
 - Dialogue States* 83
 - State Transition Rules* 83
 - Back Channel Feedback* 84
- 6.2 Discussion 84
- 6.3 The Problem with J.Jr. 84
 - The Sensing Problem* 85
 - The Lack of Behaviors Problem* 85
 - The Reactive-Reflective Integration Problem* 85
 - The Expansion Problem* 85

7. *Ymir: A Generative Model of Psychosocial Dialogue Skills* 89

- 7.1 Overview of Architectural Characteristics 90
- 7.2 The 6 Main Elements of Ymir 91
 - Layers* 92
 - Blackboards* 96
 - Perceptual Modules* 97
 - Decision Modules* 100
 - Representation of Behaviors* 100
 - Knowledge Bases: Content Interpretation & Response Generation* 105

- 7.3 Ymir: Summary of all Elements 107
- 7.4 A Notation System for Face-to-Face Dialogue Events 108
- 7.5 Summary 109



Ymir: An Implementation in LISP 111

- 8.1 Overview of Implementation 111
 - Simplifications* 111
 - Hardware Overview* 112
 - Software Overview* 112
 - Top-Level Loop* 113
- 8.2 Reactive Layer 113
 - Perceptual Modules* 113
 - Decision Modules* 116
- 8.3 Process Control Layer 119
 - Decision Modules* 119
 - Communication via the Content Blackboard* 119
- 8.4 Content Layer 119
 - Dialogue Knowledge Base* 119
 - The Topic Knowledge Base* 120
- 8.5 Action Scheduler 120
 - Behaviors* 120
 - Behavior Requests* 121
 - Generating Behavior Morphologies* 122
 - Motor Control in the Action Scheduler* 123
 - Motor Programs: Animation Unit* 124
- 8.6 Appendix: Logic Net 126
 - Syntax* 126
 - Logic Net: Any Alternatives?* 127

9. *Gandalf: Humanoid One* 129

- 9.1 The Gandalf Prototype: Overview 129
 - Prototype Setup* 129
- 9.2 Gandalf: Technical Description 132
 - Where do Gandalf's Control Rules Come From?* 132
 - Virtual Sensors* 132
 - Multimodal Descriptors* 134
 - Decision Modules* 134
- 9.3 Spatial Data Handling 137
 - Spaces & Positional Elements* 141
 - Directional Elements* 142
- 9.4 Prosody 143
 - Future Additions* 144
- 9.5 Topic & Dialogue Knowledge Bases 146
 - Speech Recognition* 146
 - Natural Language Parsing & Interpretation* 147
 - Multimodal Parsing & Interpretation* 148
 - Topic: The Solar System* 148
- 9.6 Action Scheduler 149
 - Behaviors* 149
 - Motor System* 149
 - Behavior Lexicon* 149
- 9.7 Examples of System Performance 149
 - Behavior Lexicon Listing* 153

10. *Ymir / Gandalf: An Evaluation in Three Parts* 157

- 10.1 Evaluating Gandalf with the Model Human Processor 157
 - Perceptual Processes* 158

	<i>Cognitive Processes</i>	159
	<i>Motor Processes</i>	160
	<i>Full-Loop Actions</i>	160
	<i>Conclusion</i>	161
10.2	Human Subjects Experiment	161
	<i>Background & Motivation</i>	162
	<i>Goals</i>	163
	<i>Experimental Design</i>	165
	<i>Results</i>	169
	<i>Discussion</i>	174
10.3	Ymir as a Foundation for Humanoid Agent research: Some Observations	175
	<i>Developing New Modules with the Multimodal Recorder</i>	175
	<i>Adding Functionality: Deictic Gesture at the Input</i>	177
	<i>Summary</i>	178

11. *Designing Humanoid Agents: Some High-Level Issues* 181

11.1	Validity Types	181
	<i>Face Validity</i>	182
	<i>Functional Validity</i>	182
	<i>Structural Validity</i>	183
11.2	Functional Validity in Humanoid Computer Characters	183
11.3	<i>What is my Agent?</i>	184
11.4	<i>The Distributed Agent</i>	186
	<i>Where is my Agent?</i>	186
	<i>Wristcomputer Humanoids</i>	187
	<i>A Comparison to Teleoperation</i>	188
11.5	Conclusion	190

12. *Conclusions & Future Work* 193

- 12.1 High-Level Issues 193
- 12.2 The Goals of Bridging Between Sensory Input and Action Generation 194
 - Continuous Input and Output Over Multiple Modes* 194
 - Coordination of Actions on Multiple Levels* 195
 - Lack of Behaviors* 195
 - The Natural Language Problem* 196
 - The Expansion Problem* 196
 - Goals: Conclusion* 197
- 12.3 Inherent Limitations 197
 - Reactive-Reflective Distinction* 197
 - Communication Between Layers* 197
 - Behaviors and Action Generation* 198
- 12.4 Extending Ymir/Gandalf 198
 - Where Are We Now? Current Status* 199
 - Multiple Turns for Single Utterances* 199
 - Dialogue Process Directives* 200
 - Emotional Simulation* 200
 - Spatial Sensors & their Link to Spatial Knowledge* 200
 - Dialogue History* 201
 - Advanced Gesture Recognition & Multimodal Event Representation* 201
 - Multi-Participant Conversation* 202

A1. *Character Animation* 203

- A1.1 Background, Motivation, Goals 204
- A1.2 ToonFace Architecture 205
- A1.3 The ToonFace Coding Scheme: A Comparison to FACS 210
- A1.4 Future Enhancements 211

A2. *System Specifications* 213

A2.1 Hardware 213

A2.2 Software 214

A3. *Questionnaires & Scoring* 215

A3.1 Scoring 215

A3.2 Instructions for Subjects 215

A3.3 Evaluation Questionnaire 216

A3.4 Prior Beliefs Questionnaire 219

References 223