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

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 Warshawsky 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 Bebba 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. Gu›bergsson, 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