Humanoid Robots: A New Kind of Tool

Bryan Adams, Cynthia Breazeal, Rodney A. Brooks, and Brian Scassellati, MIT Artificial Intelligence Laboratory

ASIDE FROM THEIR TRADITIONAL ROLES, HUMANOID ROBOTS CAN BE USED TO EXPLORE THEORIES OF HUMAN INTELLIGENCE. THE AUTHORS DISCUSS THEIR PROJECT AIMED AT DEVELOPING ROBOTS THAT CAN BEHAVE LIKE AND INTERACT WITH HUMANS.

In his 1923 play *R.U.R.: Rossum s Universal Robots*, Karel Capek coined *robot* as a derivative of the Czech *robota* (forced labor). Limited to work too tedious or dangerous for humans, today s robots weld parts on assembly lines, inspect nuclear plants, and explore other planets. Generally, robots are still far from achieving their fictional counterparts intelligence and flexibility.

Humanoid robotics labs worldwide are working on creating robots that are one step closer to science fiction s androids. Building a humanlike robot is a formidable engineering task requiring a combination of mechanical, electrical, and software engineering; computer architecture; and real-time control. In 1993, we began a project aimed at constructing a humanoid robot for use in exploring theories of human intelligence.[1,2] In addition to the relevant engineering, computer architecture, and real-time-control issues, we ve had to address issues particular to integrated systems: What types of sensors should we use, and how should the robot interpret the data? How can the robot act deliberately to achieve a task and remain responsive to the environment? How can the system adapt to changing conditions and learn new tasks? Each humanoid robotics lab must address many of the same motor-control, perception, and machine-learning problems.

The principles behind our methodology

The real divergence between groups stems from radically different research agendas and underlying assumptions. At the MIT AI Lab, three basic principles guide our research

- We design humanoid robots to act autonomously and safely, without human control or supervision, in natural work environments and to interact with people. We do not design them as solutions for specific robotic needs (as with welding robots on assembly lines). Our goal is to build robots that function in many different real-world environments in essentially the same way.
- Social robots must be able to detect and understand natural human cues the low-level social conventions that people understand and use everyday, such as head nods or eye contact so that anyone can interact with them without special training or instruction. They must also be able to employ those conventions to perform an interactive exchange. The necessity of these abilities influences the robots control-system design and physical embodiment.
- Robotics offers a unique tool for testing models drawn from developmental psychology and cognitive science. We hope not only to create robots inspired by biological capabilities, but also to help shape and refine our understanding of those capabilities. By applying a theory to a real system, we test the hypotheses and can more easily judge them on their content and coverage.

Autonomous robots in a human environment

Unlike industrial robots that operate in a fixed environment on a small range of stimuli, our robots must operate flexibly under various environmental conditions and for a wide range of tasks. Because we require the system to operate without human control, we must address research issues such as behavior selection and attention. Such autonomy often represents a trade-off between performance on particular tasks and generality in dealing with a broader range of stimuli. However, we believe that building autonomous systems provides robustness and flexibility that task-specific systems can never achieve.

Requiring our robots to operate autonomously in a noisy, cluttered, traffic-filled workspace alongside human counterparts forces us to build systems that can cope with natural-environment complexities. Although these environments are not nearly as hostile as those planetary explorers face, they are also not tailored to the robot. In addition to being safe for human interaction and recognizing and responding to social cues, our robots must be able to learn from human demonstration.

The implementation of our robots reflects these research principles. For example, Cog (see Figure 1) began as a 14-degrees-of-freedom (DOF) upper torso with one arm and a rudimentary visual system. In this first incarnation, we implemented multimodal behavior systems, such as reaching for a visual target. Now, Cog features two six-DOF arms, a seven-DOF head, three torso joints, and much richer sensory systems. Each eye has one camera with a narrow field of view for high-resolution vision and one with a wide field of view for peripheral vision, giving the robot a binocular, variable-resolution view of its environment. An inertial system lets the robot coordinate motor responses more reliably. Strain gauges measure the output torque on each arm joint, and potentiometers measure position. Two microphones provide auditory input, and various limit switches, pressure sensors, and thermal sensors provide other proprioceptive inputs.

Figure 1. Our upper-torso development platform, Cog, has 22 degrees of freedom that we specifically designed to emulate human movement as closely as possible.

The robot also embodies our principle of safe interaction on two levels. First, we connected the motors on the arms to the joints in series with a torsional spring.[3] In addition to providing gearbox protection and eliminating high-frequency collision vibrations, the spring s compliance provides a physical measure of safety for people interacting with the arms. Second, a spring law, in series with a low-gain force control loop, causes each joint to behave as if controlled by a low-frequency spring system (soft springs and large masses). Such control lets the arms move smoothly from posture to posture with a relatively slow command rate, and lets them deflect out of obstacles way instead of dangerously forcing through them, allowing safe and natural interaction. (For discussion of Kismet, another robot optimized for human interaction, see Social Constraints on Animate Vision, by Cynthia Breazeal and her colleagues, in this issue.)

Interacting socially with humans

Because our robots must exist in a human environment, social interaction is an important facet of our research. Building social skills into our robots provides not only a natural means of human—machine interaction but also a mechanism for bootstrapping more complex behavior. Humans serve both as models the robot can emulate and instructors that help shape the robot s behavior. Our current work focuses on four social-interaction aspects: an emotional model for regulating social dynamics, shared attention as a means for identifying saliency, acquiring feedback through vocal prosody, and learning through imitation.

Regulating social dynamics through an emotional model. One critical component for a socially intelligent robot is an emotional model that understands and manipulates its environment. A robot requires two skills to learn from such a model. First is the ability to acquire social input to understand the relevant clues humans provide about their emotional state that can help it understand any given interaction s dynamics. Second is the ability to manipulate the environment to express its own emotional state in such a way that it can affect social-interaction dynamics. For example, if the robot is observing an instructor demonstrating a task, but the instructor is moving too quickly for the robot to follow, the robot can display a confused expression. The instructor naturally interprets this display as a signal to slow down. In this way, the robot can influence the instruction s rate and quality. Our current architecture incorporates a motivation model that encompasses these exchange types (see Figure 2).

Figure 2. A generic control architecture under development for use on two of our humanoid robots. Under each large system, we list components that we either have implemented or are developing. Also, many skills reside in the interfaces between these modules, such as learning visual-motor skills and regulating attention preferences based on motivational state. We do not list machine learning techniques—an integral part of these individual systems—individually here.

Identifying saliency through shared attention. Another important requirement for a robot to participate in social situations is to understand the basics of shared attention as expressed by gaze direction, pointing, and other gestures. One difficulty in enabling a machine to learn from an instructor is ensuring the machine and instructor both attend to the same object to understand where new information should be applied. In other words, the student must know which scene parts are relevant to the lesson at hand. Human students use various social cues from the instructor for directing their attention; linguistic determiners (such as this or that), gestural cues (such as pointing or eye direction), and postural cues (such as proximity) can all direct attention to specific objects and resolve this problem. We are implementing systems that can recognize the social cues that relate to shared attention and that can respond appropriately based on the social context.

Acquiring feedback through speech prosody. Participating in vocal exchange is important for many social interactions. Other robotic auditory systems have focused on recognition of a small hardwired command vocabulary. Our research has focused on understanding vocal patterns more fundamentally. We are implementing an auditory system to let our robots recognize vocal affirmation, prohibition, and attentional bids. By doing so, the robot will obtain natural social feedback on which actions it has and has not executed successfully. Prosodic speech patterns (including pitch, tempo, and vocal tone) might be universal; infants have demonstrated the ability to recognize praise, prohibition, and attentional bids even in unfamiliar languages.

Learning through imitation. Humans acquire new skills and new goals through imitation. Imitation can also be a natural mechanism for a robot to acquire new skills and goals.[4] Consider this example:

The robot is observing a person opening a glass jar. The person approaches the robot and places the jar on a table near the robot. The person rubs his hands together and then sets himself to removing the lid from the jar. He grasps the glass jar in one hand and the lid in the other and begins to unscrew the lid by turning it counter-clockwise. While he is opening the jar, he pauses to wipe his brow, and glances at the robot to see what it is doing. He then resumes opening the jar. The robot then attempts to imitate the action.

Although classical machine learning addresses some issues this situation raises, building a system that can learn from this type of interaction requires a focus on additional research questions. Which parts of the action to be imitated are important (such as turning the lid counter-clockwise), and which aren t (such as wiping your brow)? Once the action has been performed, how does the robot evaluate the performance? How can the robot abstract the knowledge gained from this experience and apply it to a similar situation? These questions require knowledge about not only the physical but also the social environment.

Constructing and testing human-intelligence theories

In our research, not only do we draw inspiration from biological models for our mechanical designs and software architectures, we also attempt to use our implementations of these models to test and validate the original hypotheses. Just as computer simulations of neural nets have been used to explore and refine models from neuroscience, we can use humanoid robots to investigate and validate models from cognitive science and behavioral science. We have used the following four examples of biological models in our research.

Development of reaching and grasping. Infants pass through a sequence of stages in learning hand-eye coordination.[5] We have implemented a system for reaching to a visual target that follows this biological model.[6] Unlike standard kinematic manipulation techniques, this system is completely self-trained and uses no fixed model of either the robot or the environment.

Similar to the progression observed in infants, we first trained Cog to orient visually to an interesting object. The robot moved its eyes to acquire the target and then oriented its head and neck to face the target. We then trained the robot to reach for the target by interpolating between a set of postural primitives that mimic the responses of spinal neurons identified in frogs and rats.[7] After a few hours of unsupervised training, the robot executed an effective reach to the visual target.

Several interesting outcomes resulted from this implementation. From a computer science perspective, the two-step training process was computationally simpler. Rather than attempting to map the visual-stimulus location s two dimensions to the nine DOF necessary to orient and reach for an object, the training focused on learning two simpler mappings that could be chained together to produce the desired behavior. Furthermore, Cog learned the second mapping (between eye position and the postural primitives) without supervision. This was possible because the mapping between stimulus location and eye position provided a reliable error signal (Figure 3). From a biological standpoint, this implementation uncovered a limitation in the postural primitive theory. Although the model described how to interpolate between postures in the initial workspace, no mechanism for extrapolating to postures outside the initial workspace existed.

Figure 3. Reaching to a visual target. Once the robot has oriented to a stimulus, a ballistic mapping computes the arm commands necessary to reach for that stimulus. The robot observes its own arm's motion. It then uses the same mapping that it uses for orientation to produce an error signal it can use to train the ballistic map.

Rhythmic movements. Kiyotoshi Matsuoka[8] describes a model of spinal cord neurons that produce rhythmic motion. We have implemented this model to generate repetitive arm motions, such as turning a crank.[9] Two simulated neurons with mutually inhibitory connections drive each arm joint, as Figure 4 shows. The oscillators take proprioceptive input from the joint and continuously modulate the equilibrium point of that joint s virtual spring. The interaction of the oscillator dynamics at each joint and the arm s physical dynamics determines the overall arm motion.

Figure 4. Neural oscillators. The oscillators attached to each joint comprise a pair of mutually inhibiting neurons. Black circles represent inhibitory connections; open white circles are excitatory. The final output is a linear combination of the neurons' individual outputs.

This implementation validated Matsuoka s model on various real-world tasks and provided some engineering benefits. First, the oscillators require no kinematic model of the arm or dynamic model of the system. No a priori knowledge was required about either the arm or the environment. Second, the oscillators were able to tune to a wide task range, such as turning a crank, playing with a Slinky, sawing a wood block, and swinging a pendulum, all without any change in the control system configuration. Third, the system was extremely tolerant to perturbation. Not only could we stop and start it with a very short transient period (usually less than one cycle), but we could also attach large masses to the arm and the system would quickly compensate for the change. Finally, the input to the oscillators could come from other modalities. One example was using an auditory input that let the robot drum along with a human drummer.

Visual search and attention. We have implemented Jeremy Wolfe s model of human visual search and attention,[10] combining low-level feature detectors for visual motion, innate perceptual classifiers (such as face detectors), color saliency, and depth segmentation with a motivational and behavioral model (see Figure 5). This attention system lets the robot selectively direct computational resources and exploratory behaviors toward objects in the environment that have inherent or contextual saliency.

Figure 5. Attention system overview. Various visual-feature detectors (color, motion, and face detectors) combine with a habituation function to produce an attention activation map. The attention process influences eye control and the robot's internal motivational and behavioral state, which in turn influence the weighted feature-map combination. We captured the images during a behavioral trial session.

This implementation has let us demonstrate preferential looking based both on top-down task constraints and opportunistic use of low-level features.[11] For example, if the robot is searching for ocial contact, the motivation system increases the weight of the face-detector feature. This produces a preference for looking at faces. However, if a very interesting nonface object appeared, the object s low-level properties would be sufficient to attract the robot s attention. We are incorporating saliency cues based on the model s focus of attention into this attention model. We were also able to devise a simple mechanism for incorporating habituation effects into Wolfe s model. By treating time-decayed Gaussian fields as an additional low-level feature, the robot will habituate to stimuli that are currently receiving attentional resources.

Shared attention and theory of mind. One critical milestone in a child s development is the recognition that others have beliefs, desires, and perceptions independent of the child s. The abilities to recognize what another person can see, know that another person maintains a false belief, and recognize that another person likes games differing from those the child enjoys are all part of this developmental chain. Furthermore, the ability to recognize yourself in the mirror, the ability to ground words in perceptual experiences, and the skills involved in creative and imaginative play might also be related to this developmental advance. We are implementing a model of social-skill development that accounts for both normal development and the developmental disorders associated with autism. We have implemented systems that can detect faces and eyes in unconstrained visual environments and are working on detecting eye contact.

Although this work is still preliminary, we believe that implementing a developmental model on a robot will allow detailed and controlled manipulations of the model while maintaining the same testing environment and methodology used on human subjects. Researchers can vary internal model parameters systematically as they evaluate the effects of different environmental conditions on each step of development. Because the robot brings the model into the same environment as a human subject, researchers can use similar evaluation criteria (whether subjective measurements from observers or quantitative measurements such as reaction time or accuracy). Also, researchers can subject a robot to testing that s potentially hazardous, costly, or unethical to conduct on humans.

*A*lthough scientific research usually takes credit as the inspiration for science fiction, it s possible that with AI and robotics, fiction led the way. However, over the past 10 years, humanoid robotics has become the focus of many research groups, conferences, and special issues. While outpacing the imagination of science-fiction writers might be difficult, our work does indicate one possible future. Robots will be able to interact with humans in humanlike ways, and people will find this normal and natural. At the same time, we will continue to learn more about the nature of our own intelligence by building these systems.

Acknowledgemnts

This work was supported by ONR and DARPA under MURI N00014-95-1-0600 and by DARPA under contract DABT 63-99-1-0012.

References

1. R.A. Brooks et al., Alternative Essences of Intelligence, *Proc. 15th Nat l Conf. Artificial Intelligence* (AAAI 98) and *10th Conf. Innovative Applications of Artificial Intelligence* (IAAI 98), AAAI Press, Menlo Park, Calif., 1998, pp. 961—968.
2. R.A. Brooks et al., The Cog Project: Building a Humanoid Robot, *Computation for Metaphors, Analogy and Agents*, C. Nehaniv, ed., *Springer Lecture Notes in Artificial Intelligence*, Vol. 1562, Springer-Verlag, Berlin, 1998.
3. G.A. Pratt and M.M. Williamson, Series Elastic Actuators, *Proc. IEEE/RSJ Int l Conf. Intelligent Robots and Systems* (IROS 95), Vol. 1, IEEE Computer Soc. Press, Los Alamitos, Calif., 1995, pp. 399—406.
4. C. Breazeal and B. Scassellati, Challenges in Building Robot s That Imitate People, to be published in *Imitation In Animals and Artifacts*, K. Dautenhahn and C. Nehaniv, eds., MIT Press, Cambridge, Mass., 2000.
5. A. Diamond, Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of Inhibitory Control in Reaching, *The Development and Neural Bases of Higher Cognitive Functions*, New York Academy of Sciences, New York, 1990, pp. 637—676.
6. M.J. Marjanovic, B. Scassellati, and M.M. Williamson, Self-Taught Visually Guided Pointing for a Humanoid Robot, *From Animals to Animats 4: Proc. Fourth Int l Conf. Simulation of Adaptive Behavior* (SAB 96), MIT Press, Cambridge, Mass, 1996, pp. 35—44.
7. S.F. Giszter, F.A. Mussa-Ivaldi, and E. Bizzi, Convergent Force Fields Organized in the Frog s Sp inal Cord, *J. Neuroscience*, Vol. 13, No. 2, February 1993, pp. 467—491.
8. K. Matsuoka, Sustained Oscillations Generated by Mutually Inhibiting Neurons with Adaption, *Biological Cybernetics*, Vol. 52, 1985, pp. 367—376.
9. M. Williamson, *Robot Arm Control Exploiting Natural Dynamics*, doctoral thesis, Massachusetts Institute of Technology, Dept. Electrical Eng. and Computer Science, Cambridge, Mass., 1999.
10. J. Wolfe, Guided Search 2.0: A Revised Model of Visual Search, *Psychonomic Bull. and Rev.*, Vol. 1, No. 2, June 1994, pp. 202—238.
11. C. Breazeal and B. Scassellati, A Context-Dependent Attention System for a Social Robot, *Proc. 16th Int l Joint Conf. Artificial Intelligence* (IJCAI 99), Morgan Kaufmann, San Francisco, 1999, pp. 1146—1153.

Bryan Adams is completing his ME with Rodney Brooks Humanoid Robotics group and is interested in theories of intelligent control for humanoid arms. He received his BS in electrical engineering and computer science from MIT. Contact him at the MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139; bpadams@ai.mit.edu; www.ai.mit.edu/people/bpadams.

Cynthia Breazeal received her ScD from the MIT Artificial Intelligence Laboratory. Her interests focus on humanlike robots that can interact in natural, social ways with humans. She received her BS in electrical and computer engineering from the University of Calif., Santa Barbara, and her MS in electrical engineering and computer science from MIT. Contact her at the MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139; cynthia@ai.mit.edu; www.ai.mit.edu/people/cynthia/cynthia.html.

Rodney A. Brooks is the Director of the MIT Artificial Intelligence Laboratory and the Fujitsu Professor of Computer Science and Engineering. His research interests include robotics, computer vision, and architectures for intelligence. He received his PhD in computer science from Stanford. He is an IEEE member and a fellow of both the American Association for Artificial Intelligence and American Association for the Advancement of Science. Contact him at the MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139; brooks@ai.mit.edu; www.ai.mit.edu/people/brooks.

Brian Scassellati is completing his PhD with Rodney Brooks at the MIT Artificial Intelligence Laboratory. His work is strongly grounded in theories of how the human mind develops, and he is interested in robotics as a tool for evaluating models from biological sciences. He received a BS in computer science, a BS in brain and cognitive science, and an ME in electrical engineering and computer science from MIT. Contact him at the MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139; scaz@ai.mit.edu; www.ai.mit.edu/people/scaz.