Humanoid Robots: A New Kind of Tool

Bryan Adams, Cynthia Breazeal, Rodney Brooks, Brian Scassellati

MIT Artificial Intelligence Laboratory

545 Technology Square

Cambridge, MA 02139

USA

{bpadams, cynthia, brooks, scaz}@ai.mit.edu

Abstract

In 1993 our group at the MIT Artificial Intelligence Laboratory began a humanoid robotics project aimed at constructing a robot for use in exploring theories of human intelligence.[1 2] In this article, we will describe three aspects of our research methodology that distinguish our work from other humanoid projects. First, our humanoid robots are designed to act autonomously and safely in natural workspaces with people. Second, our robots are designed to interact socially with people by exploiting natural human social cues. Third, we believe that robotics offers a unique tool for testing models of human intelligence drawn from developmental psychology and cognitive science.

Keywords

Humanoid, robotics, autonomous, embodied, social, attention, imitation.

IEEE Intelligent Systems

Introduction

While scientific research usually takes credit as the inspiration for science fiction, in the case of AI and robotics, it is possible that fiction led the way for science. The term robot was coined in a 1923 play by the Capek Brothers, entitled *RUR* (Rossum's Universal Robots), as a derivative of the Czech *robota* which means "forced labor." Today s robots weld parts on assembly lines, inspect nuclear power plants, and explore the surface of other planets. They are limited to forced labor that is either too tedious or too dangerous for humans. Generally speaking, robots of today are still far from achieving the intelligence and flexibility of their fictional counterparts.

Today, humanoid robotics labs across the globe are working on creating a new set of robots that take us one step closer to the androids of science fiction. Building a human-like robot is a formidable engineering task that requires a combination of mechanical engineering, electrical engineering, computer architecture, real-time control, and software engineering. Research issues from each of these fields as well as issues particular to integrated systems and robotics must be addressed to build a robot: What types of sensors should be used and how should the data be interpreted? How can the motors be controlled 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 problems of motor control, perception, and machine learning. The real divergence between groups stems from radically different research agendas and underlying assumptions. At the MIT Artificial Intelligence lab, our research is guided by three basic principles.

First, our humanoid robots are designed to act autonomously and safely in natural workspaces with people. Our robots are not designed as a solution to a specific robotic need (as the welding robot on an assembly line would be). Instead, they are designed to exist and interact with the world in a way similar to how a typical person would. As opposed to a robot that operates in an environment engineered specifically for the robot, we engineer our robots to operate in typical, everyday environments. Our goal is to build robots that function in many different real-world environments in essentially the same way.

Second, our robots are designed to interact socially with people by exploiting natural human social cues. Instead of asking people to interact with our robots in a specific, predetermined way, we try to engineer our robots to interact with people in the same ways that people interact with each other. This allows anyone to interact with the robot without requiring special training or instruction. A social robot requires the ability to detect and understand the low-level social conventions that people understand and use in everyday interactions, such as head nods or eye contact. It also requires the ability to then put the conventions to work on the behalf of the robot to complete the interactive exchange. This influences the design of both the control system for the robots and the physical embodiment of the robots themselves.

Third, we believe that robotics offers a unique tool for testing models drawn from developmental psychology and cognitive science. We hope not only to produce robots that are inspired by biological capabilities, but also to help shape and refine our understanding of those capabilities. By bringing a theory to bear on a real system, the proposed hypotheses are tested in the real world and can be more easily judged on their content and coverage.

In this paper, we will take each of these guidelines and examine it more closely in the light of the robots that we have designed and built, the systems that have already been constructed, and our plans for future development.

1 Autonomous Robots in a Human Environment

Our research focuses on building autonomous robots that are not under human control or supervision. Unlike industrial robots that operate in a fixed environment on a small range of stimuli, our robots must operate flexibly under a variety of 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. Autonomy of this kind 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.

In addition to being autonomous, we require that our robots function in the human environment. The robot must operate in a noisy, cluttered, traffic-filled workspace alongside human counterparts. This requirement forces us to build systems that can cope with the complexities of natural environments. While these environments are not nearly as hostile as those faced by planetary explorers, they are also not tailored to the robot. These requirements force us to construct robots that are safe for human interaction and that address research issues such as recognizing and responding to social cues and learning from human demonstration.

The implementation of our robots reflects these research principles. Cog (Figure 1) began as a 14 degree-of-freedom upper torso with one arm and a rudimentary visual system. In this first incarnation, multimodal behavior systems, such as reaching for a visual target, were implemented. Currently, Cog features two six degree-of-freedom arms, a seven degree-of-freedom head, three torso joints, and a much richer array of sensors. 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 allows the robot to coordinate motor responses more reliably. Strain gauges measure the output torque on each of the joints in the arm and potentiometers provide an accurate measure of the position. Two microphones provide auditory input, and a variety of limit switches, pressure sensors, and thermal sensors provide other proprioceptive inputs.

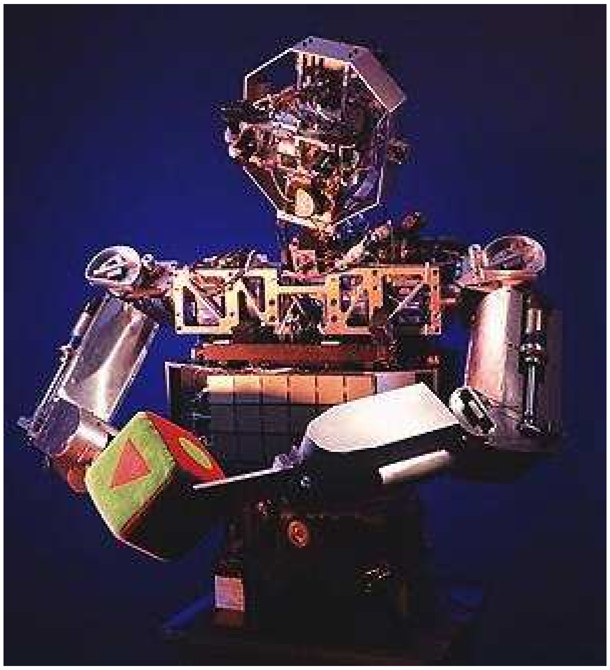

Figure 1: Our upper-torso development platform, Cog, has twenty-two degrees of freedom that are specifically designed to emulate human movement as closely as possible.

The robot also embodies our principle of safety of interaction on two levels. First, the motors on the arms are connected to the joints in series with a torsional spring.[3] In addition to providing protection to the gearbox and eliminating high-frequency vibrations from collision, the compliance of the spring provides a physical measure of safety for those interacting with the arms. Second, a spring law, in series with a low-gain force control loop, causes each joint to behave as if it were controlled by a low-frequency spring (soft springs and large masses). This type of control allows the arms to move smoothly from posture to posture with a relatively slow command rate, but also causes them to deflect out of the way of obstacles instead of dangerously forcing through them, allowing for safe and natural interaction.

Kismet (Figure 2) began as an active vision platform, using only a pair of eyes to interact with the world. Additional facial features were added to provide more expressive capabilities. The robot s internal state and perceived visual stimuli combine to produce a three-dimensional measurement of the robot s emotional state. Primitive facial expressions are blended together based on this emotional state to produce a continuously varying facial expression and posture.[4] More recent research incorporated an auditory system and a speech synthesizer to allow the robot to participate in verbal interactions with its caregiver.

Figure 2: Kismet, the emotional/visual development platform, uses twenty-one degrees of freedom to express its emotional state.

2 Interacting Socially with Humans

Because our robots exist autonomously in a human environment, engaging in 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 that the robot can emulate and as instructors that help to shape the robot s behavior. Our current work focuses on four aspects of social interaction: an emotional model for regulating social dynamics, shared attention as a means for identifying saliency, acquiring feedback through vocal prosody, and learning through imitation.

2.1 Regulating social dynamics through an emotional model. One critical component for a socially intelligent robot is an emotional model that understands and manipulates the environment around it. This requires two skills. The first is the ability to acquire social input; to understand the relevant clues that humans provide about their emotional state that can be helpful in understanding the dynamics of any given interaction. The second is the ability to manipulate the environment; for a robot to express its own emotional state in such a way that it can affect the dynamics of social interaction. 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 an expression of confusion. This display is naturally interpreted by the instructor as a signal to slow down. In this way, the robot can influence the rate and quality of the instruction. Our current architecture incorporates a model of motivation that encompasses these types of exchanges (Figure 3).

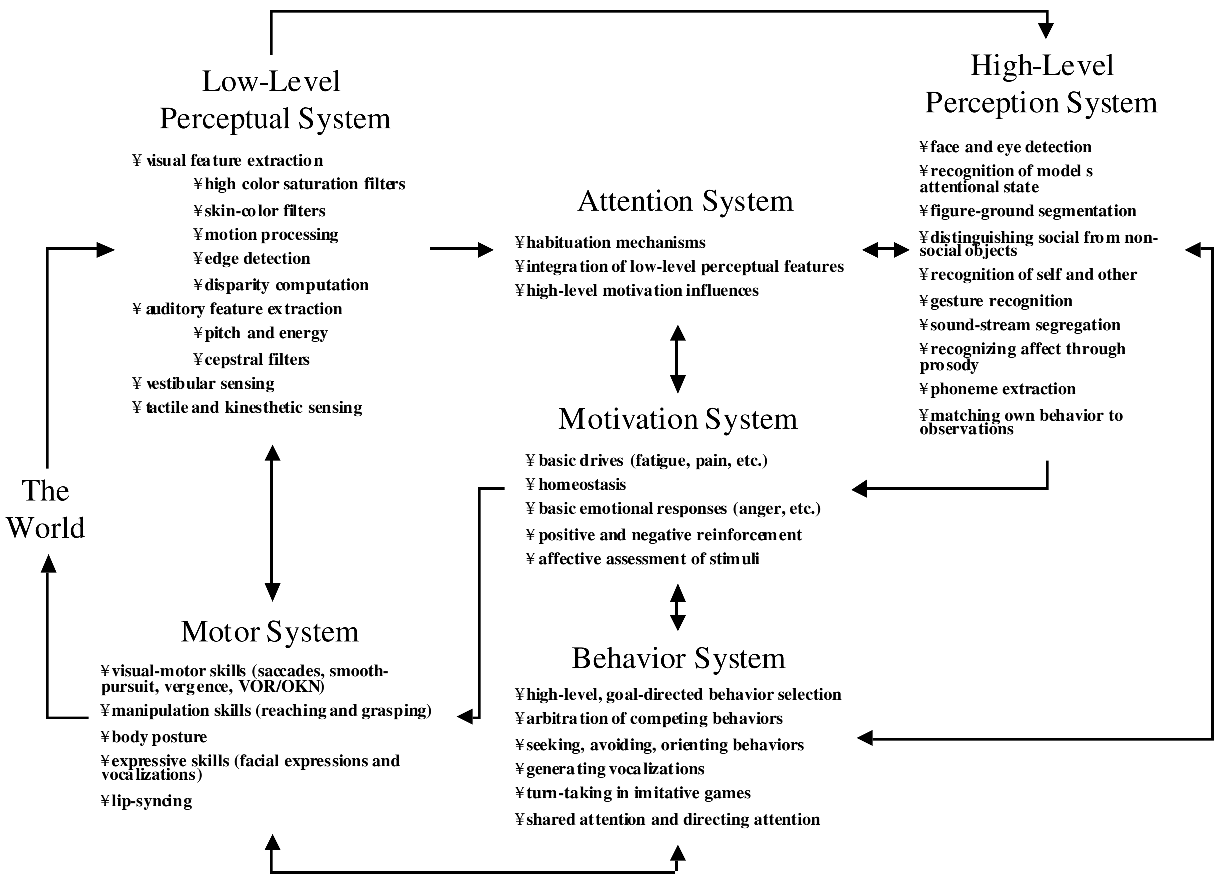

Figure 3: A generic control architecture under development for use on our humanoid robots Cog and Kismet. Under each large system, we have listed components that have either been implemented or are currently under development. There are also many skills that reside in the interfaces between these modules, such as learning visual-motor skills and regulating attention

preferences based on motivational state. Machine learning techniques are an integral part of each of these individual systems, but are not listed individually here.

2.2 Determining saliency through shared attention. Another important component 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 that the student and the instructor are both attending to the same object in order to understand where new information should be applied. In other words, the student must know which parts of the scene are relevant to the lesson at hand. Human students use a variety of 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 currently engaged in implementing systems that can recognize the social cues that relate to shared attention and that can respond appropriately based on the social context.

2.3 Social feedback through speech prosody. Participating in vocal exchange is an important part of many social interactions. Other robotic auditory systems have focused on recognition of a small vocabulary of hard-wired commands. Our research has focused on understanding speech patterns in a more fundamental way. We are currently implementing an auditory system to enable our robots to recognize vocal affirmation, prohibition, and attentional bids while interacting with a human. By doing so, the robot will obtain natural social feedback on which of its actions have been successfully executed and which have not. Prosodic patterns of speech (including pitch, tempo, and tone of voice) may be universal, as infants have demonstrated the ability to recognize praise, prohibition and attentional bids even in unfamiliar languages.

2.4 Learning through imitation. Humans acquire new skills and new goals through imitation. Imitation can also be a natural mechanism for a robot in human environments to acquire new skills and goals.[5] Consider the following 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.

While classical machine learning addresses some of the issues raised by this situation, building a system that can learn from this type of interaction requires a focus on additional research questions. What parts of the task to be imitated are important (like turning the lid counter-clockwise) and which parts are unimportant (like wiping your brow)? Given some sort of behavior-response, how does the robot evaluate its performance? How can the robot abstract the knowledge gained from this experience and apply it to a similar situation? These questions require knowledge not only about the physical environment, but about the social environment as well.

3 Constructing and Testing Theories of Human Intelligence

A major focus of our group is not only on constructing intelligent machines, but also on using those machines as a means for testing ideas about the nature of human intelligence. In our research, not only do we draw inspiration from biological models for our mechanical designs and software architectures, but 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, humanoid robots can be used to investigate and validate models from cognitive science and behavioral science. The following are four examples of biological models that have been used in our research.

3.1 A model of the development of reaching behavior based on infant studies.

Infants pass through a sequence of stages in learning hand-eye coordination.[6] We have implemented a system for reaching to a visual target that follows this biological model.[7] Unlike standard kinematic techniques for manipulation, this system is completely self-trained and uses no fixed model of either the robot or the environment.

Similar to the progression of infants, we first trained the robot 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. The robot was then trained to reach for the target by interpolating between a set of postural primitives that mimic the responses of spinal neurons that have been identified in the frog and rat.[8] Over the course of a few hours of unsupervised training, the robot was able to execute an effective reach to the visual target.

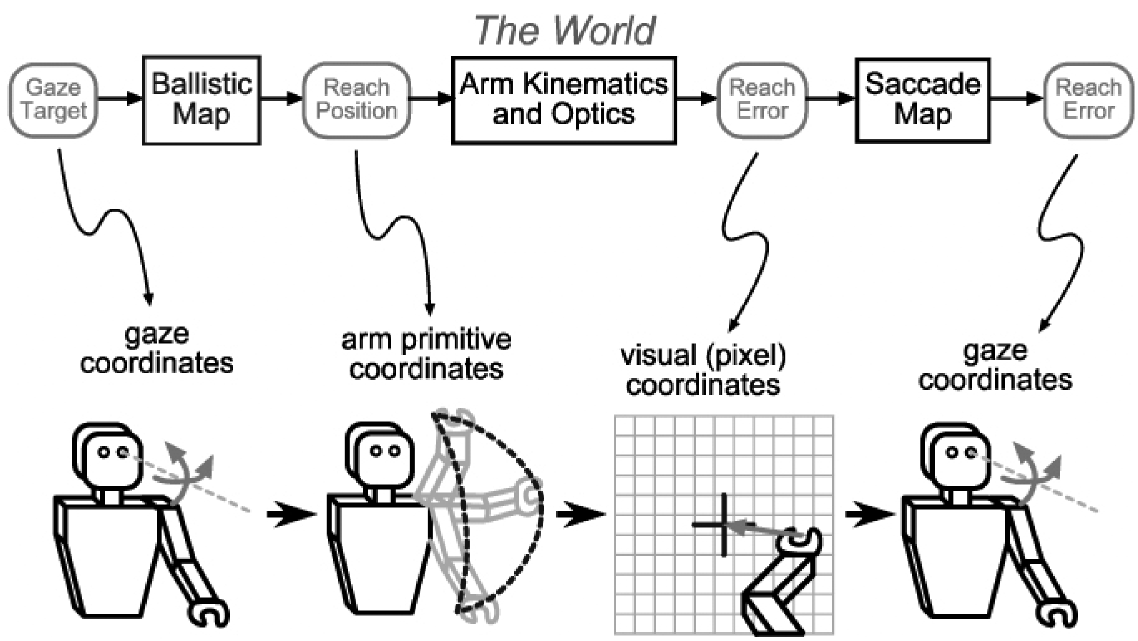

Figure 4: Reaching to a visual target. Once the robot has oriented to a stimulus, a ballistic mapping computes the arm commands necessary to reach for that stimulus. The robot observes the motion of its own arm, and then uses the same mapping that is used for orientation to produce an error signal that can be used to train the ballistic map.

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 two-dimensions of the location of the visual stimulus to the nine degrees of freedom 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, training the second mapping (between eye position and the postural primitives) could be accomplished without supervision because the mapping between stimulus location and eye position could provide a reliable error signal (Figure 4). From a biological standpoint, this implementation uncovered a limitation in the postural primitive theory. This model had no mechanism for representing movements or spatial positions outside the workspace defined by the set of initial primitive postures. Although the model described how to interpolate between postures within the initial workspace, there was no mechanism for extrapolating to postures outside the initial workspace.

3.2 A model of rhythmic motor skills based on neural oscillator circuits in the spinal cord. Matsuoka[9] 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.[10] Two simulated neurons with mutually inhibitory connections drive each arm joint, as shown in Figure 5. The oscillators take proprioceptive input from the joint and continuously modulate the equilibrium point of that joint s virtual spring (see section 1.3). The interaction of the oscillator dynamics at each joint and the physical dynamics of the arm determines the overall arm motion.

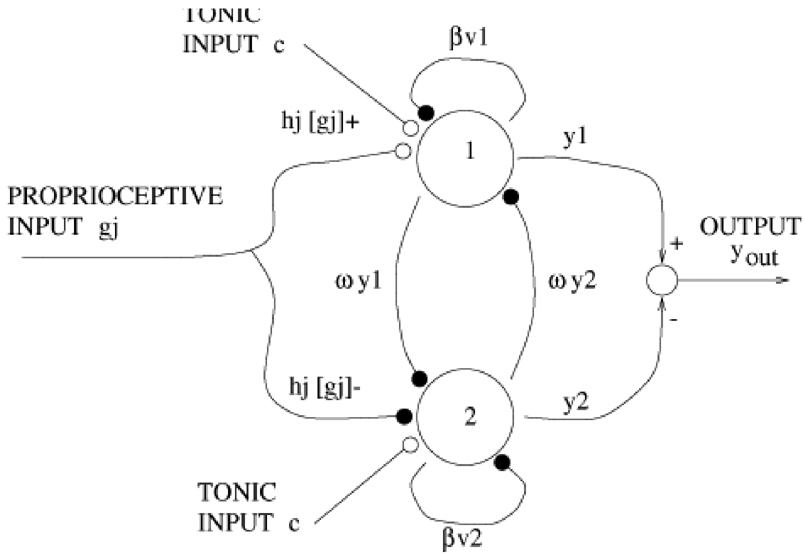

Figure 5: (Neural Oscillators) The oscillators attached to each joint are made up of a pair of mutually inhibiting neurons. Black circles represent inhibitory connections while open white circles are excitatory. The final output is a linear combination of the outputs of each of the neurons.

This implementation validated Matsuoka s model on a variety of real-world tasks and provided a number of 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 range of tasks such as turning a crank, playing with a slinky toy, sawing a block of wood, and swinging a pendulum, all without any change in the configuration of the control system. Third, the system was extremely tolerant to perturbation. Not only could the system be stopped and started with a very short transient period (usually less than one cycle), but also large masses could be attached to the arm and the system was able to quickly attenuate the change. Finally, the input to the oscillators could come from other modalities. One example was using an auditory input that allowed the robot to drum along with a human drummer.

3.3 A model of visual search and attention. We have implemented Wolfe s model of human visual search and attention[11] that combines information from low-level features with high-level motivational influences. Our implementation combines 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 (Figure 6). This attention system allows the robot to selectively direct computational resources and exploratory behaviors toward objects in the environment that have inherent or contextual saliency.

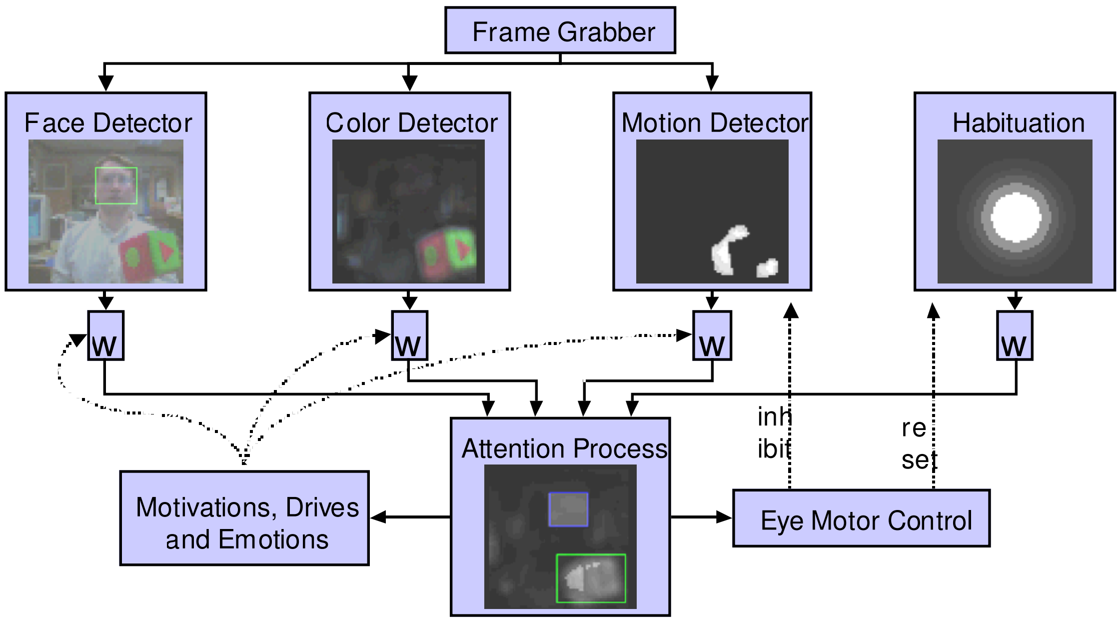

Figure 6: Overview of the attention system. A variety of visual feature detectors (color, motion, and face detectors) combine with a habituation function to produce an attention activation map. The attention process influences eye control and the robot's internal motivational and behavioral state, which in turn influence the weighted combination of the feature maps. Displayed images were captured during a behavioral trial session.

This implementation has allowed us to demonstrate preferential looking based both on top-down task constraints and opportunistic use of low-level features.[12] For example, if the robot is searching for a playmate, the weight of the face detector can be increased to cause the robot to show a preference for attending to faces. However, if a very interesting non-face object were to appear, the low-level properties of the object would be sufficient to direct attention. The addition of saliency cues based on the model s focus of attention can easily be incorporated into this model of attention, but the perceptual abilities needed to obtain the focus of attention have yet to be fully developed. We were also able to suggest 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.

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

While this work is still preliminary, we believe that having an implementation of 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. Internal model parameters can be varied systematically as the effects of different environmental conditions on each step in development are evaluated. Because the robot brings the model into the same environment as a human subject, similar evaluation criteria can be used (whether subjective measurements from observers or quantitative measurements such as reaction time or accuracy). Further, a robotic model can also be subjected to controversial testing that is potentially hazardous, costly, or unethical to conduct on humans.

4 Conclusion

In the past 10 years, humanoid robotics has become the focus of many research groups, conferences, and special issues. While all humanoid projects must address many of the same fundamental problems of motor control, perception, and general architecture, our group has focused on three additional aspects. We are committed to building robots that behave like creatures in real environments and interact with people in natural ways. We believe that constructing systems that can interact socially with people will lead to simpler techniques for machine learning and human-computer interfaces. Finally, we believe that not only should humanoid robotics look to biology for inspiration, but also that humanoid robotics should serve as a tool for investigating theories of human and animal cognition.

While it may be difficult for us to outpace the imaginations of science fiction writers, our work does indicate one possible future. Robots will be able to interact with humans in human-like ways, and people will find this natural and normal.

Biographical Sketches

Bryan Adams received a S.B. degree in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1999. He has worked for Prof. Rodney Brooks Humanoid Robotics group for 2 years, and is interested in theories of intelligent control for humanoid arms.

Cynthia Breazeal received her B.Sc. degree from the University of California, Santa Barbara in Electrical and Computer Engineering in 1989, and received her M.Sc. degree in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1993 She is currently completing her Ph.D. with Prof. Rodney Brooks at the MIT Artificial Intelligence Laboratory. Her current interests focus on human-like robots that can interact in natural, social ways with humans.

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 the PhD in Computer Science from Stanford in 1981.˚ He is a member of IEEE, and a fellow of both AAAI and AAAS.

Brian Scassellati received S.B. degrees in computer science and brain and cognitive science from the Massachusetts Institute of Technology in 1994, and a Masters of Engineering degree in Electrical Engineering and Computer Science from MIT in 1995. Since then, he has been a graduate student working towards his Ph.D. with Prof. 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 utilizing robotics as a tool for evaluating models from biological sciences.

---

[1] R. A. Brooks, C. Breazeal (Ferrell), R. Irie, C. C. Kemp, M. J. Marjanovic, B. Scassellati, M. M. Williamson, Alternate Essences of Intelligence, AAAI 1998.
[2] R. A. Brooks, C. Breazeal, M. J. Marjanovic, B. Scassellati, M. M. Williamson, The Cog Project: Building a Humanoid Robot, *Computation for Metaphors, Analogy and Agents*, C. Nehaniv, ed., Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, 1998, Springer-Verlag.
[3] G. A. Pratt & M. M. Williamson, Series Elastic Actuators, Proc. of the IEEE/RSJ International Conference on Intelligent Robots and System (IROS-95), Vol. 1, Pittsburgh, PA, 1995, pp. 399-406.
[4] C. Breazeal & B. Scassellati, How to Build Robots that Make Friends and Influence People, presented at the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-99), Kyongju, Korea, 1999.
[5] C. Breazeal & B. Scassellati, B., Challenges in Building Robots That Imitate People, to appear in Imitation In Animals and Artifacts, Dautenhahn, K & Nehaniv, C., eds. MIT Press, 2000.
[6] A. Diamond, Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of Inhibitory Control in Reaching, in The Development and Neural Bases of Higher Cognitive Functions, Vol. 608, 1990, New York Academy of Sciences, pp. 637-676.
[7] M. J. Marjanovic, B. Scassellati, M. M. Williamson, Self-Taught Visually-Guided Pointing for a Humanoid Robot, presented at *SAB 1996*, Cape Cod, MA.
[8] S. F. Giszter, F. A. Mussa-Ivaldi, & E. Bizzi, Convergent Force Fields Organized in the Frog's Spinal Cord, Journal of Neuroscience, Vol. 13 (2), 1993, pp. 467-491.
[9] K. Matsuoka, Sustained oscillations generated by mutually inhibiting neurons with adaption, Biological Cybernetics, Vol. 52, 1985, pp. 367-376.
[10] M. Williamson, Robot Arm Control Exploiting Natural Dynamics, Ph.D. Thesis, 1999, Massachusetts Institute of Technology.
[11] J. Wolfe, Guided search 2.0: A revised model of visual search, Psychonomic Bulletin & Review, Vol. 192, 1994, pp. 202-238.
[12] C. Breazeal, B. Scassellati, A Context-Dependent Attention System for a Social Robot, presented at the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 1999), Stockholm, Sweden.