Prospects for Human Level Intelligence for Humanoid Robots

Rodney A. Brooks
MIT Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139, USA
brooks@ai.mit.edu

Abstract

Both direct, and evolved, behavior-based approaches to mobile robots have yielded a number of interesting demonstrations of robots that navigate, map, plan and operate in the real world. The work can best be described as attempts to emulate insect level locomotion and navigation, with very little work on behavior-based non-trivial manipulation of the world. There have been some behavior-based attempts at exploring social interactions, but these too have been modeled after the sorts of social interactions we see in insects. But thinking how to scale from all this insect level work to full human level intelligence and social interactions leads to a synthesis that is very different from that imagined in traditional Artificial Intelligence and Cognitive Science. We report on work towards that goal.

1 Why Humanoid Robots?

The implicit dream of Artificial Intelligence has been to build human level intelligence. This is one more instantiation of the human dream to build an artificial human which has been present for almost all of recorded history, and regularly appears in modern popular culture. Building a humanoid robot is the challenge *par excellence* for artificial intelligence and robotics workers. Recent progress in many fields now means that is practical to make serious attempts at this goal. These attempts will not likely be completely successful in the near term. but they will shed light on many new fundamental problem areas, and provide new directions for research towards the ultimate goal.

2 New Challenges for Humanoid Robots

In order to act like a human[1] an artificial creature with a human form needs a vastly richer set of abilities in gaining sensor information, even from a single vantage point. Some of these basic capabilities include saccading to motion, eyes responding appropriately to vestibular signals, smooth tracking, coordinating body, head, and eye motions, compensating for visual slip when the head or body moves, correlating sound localization to occulomotor coordinates, maintaining a zero disparity common fixation point between eyes, and saccading while maintaining a fixation distance. In addition it needs much more coordinated motor control over many subsystems, as it must, for instance, maintain body posture as the head and arms are moved, coordinate redundant degrees of freedom so as to maximize effective sense and work space, and protect itself from self-injury.

In using arms and hands some of the new challenges include visually guided reaching, identification of self motion, compliant interaction with the world, self protective reflexes, grasping strategies without accurate geometric models, material estimation, dynamic load balancing, and smooth on-the-fly trajectory generation.

In thinking about interacting with people some of the important issues are detecting faces, distinguishing human voices from other sounds, making eye contact, following the gaze of people, understanding where people are pointing, interpreting facial gestures, responding appropriately to breaking or making of eye contact, making eye contact to indicate a change of turn in social interactions, and understanding personal space sufficiently.

Besides this significantly increased behavioral repertoire there are also a number of key issues that have not been so obviously essential in the previous work on behavior-based robots. When considering cognitive robotics, or *cognobotics*, one must deal with the following issues:

- bodily form

[1] It is difficult to establish hard and fast criteria for what it might mean to *act like a human* — roughly we mean that the robot should act in such a way that an average (whatever that might mean) human observer would say that it is acting in a human-like manner, rather than a machine-like or alien-like manner.

- motivation
- coherence
- self adaptation
- development
- historical contingencies
- inspiration from the brain

We will examine each of these issues in a little more detail in the following paragraphs. While some of these issues have been touched upon in behavior-based robotics it seems that they are much more critical in very complex robots such as a humanoid robot.

2.1 Bodily form

In building small behavior-based robots the overall morphology of the body has not been viewed as critical. The *detailed* morphology has been viewed as very critical, as very small changes in morphology have been observed to make major changes in the behavior of a robot running any particular set of programs as the dynamics of its interaction with the environment can be greatly affected. But, there has been no particular reason to make the robots morphologically similar to any particular living creature.

In thinking about human level intelligence however, there are two sets of reasons one might build a robot with humanoid form.

If one takes seriously the arguments of Johnson (1987) and Lakoff (1987), then the form of our bodies is critical to the representations that we develop and use for both our *internal thought* and our language. If we are to build a robot with human like intelligence then it must have a human like body in order to be able to develop similar sorts of representations. However, there is a large cautionary note to accompany this particular line of reasoning. Since we can only build a very crude approximation to a human body there is a danger that the essential aspects of the human body will be totally missed. There is thus a danger of engaging in cargo-cult science, where only the broad outline form is mimicked, but none of the internal essentials are there at all.

A second reason for building a humanoid form robot stands on firmer ground. An important aspect of being human is interaction with other humans. For a human-level intelligent robot to gain experience in interacting with humans it needs a large number of interactions. If the robot has humanoid form then it will be both easy and natural for humans to interact with it in a human like way. In fact it has been our observation that with just a very few human-like cues from a humanoid robot. people naturally fall into the pattern of interacting with it as if it were a human. Thus we can get a large source of dynamic interaction examples for the robot to participate in. These examples can be used with various internal and external evaluation functions to provide experiences for learning in the robot. Note that this source would not be at all possible if we simply had a disembodied human intelligence. There would be no reason for people to interact with it in a human-like way.

2.2 Motivation

Recent work in behavior-based robots has dealt with systems with only a few sensors and mostly with a single activity of navigation. The issue of the robot having any motivation to engage in different activities does not arise. The robots have some small set of navigation behaviors built in, some switching between behaviors may occur, and the robot must simply navigate about in the world. There is no choice of activities in these papers, so there needs be no mechanism to motivate the pursuit of one activity over another. The one major activity that these robots engage in, has, naturally, a special place, and so the systems are constructed so that they engage in that activity — there is no mechanism built in for them not to engage in that activity.

When we move to much more complex systems there are new considerations:

- When the system has many degrees of freedom, or actuators, it can be the case that certain activities can be engaged in using only some subsystems of the physical robot. Thus it can be the case that it may be possible for the system to engage in more than one activity simultaneously using separable sensory and mechanical subsystems.

- The system may have many possible activities that it can engage in, but these activities may conflict—i.e., if engaging in activity A it may be impossible to simultaneously engage in activity B.

Thus for more complex systems, such as with a full human level intelligence in a human form, we are faced with the problem of motivation. When a humanoid robot is placed

in a room with many artifacts around it, why should it interact with them at all? When there are people within its sensory range why should it respond to them? Unlike the mobile robot case where an implicit unitary motivation sufficed, in the case of a full humanoid robot there is the problem of a confrontation with a myriad of choices of what or who to interact with, and how that interaction should take place. The system needs to have some sort of motivations, which may vary over time both absolutely and relatively to other motivations, and these motivations must be able to reach some sort of expression in what it is that the humanoid does.

2.3 Coherence

Of course, with a complex robot, such as a humanoid robot with many degrees of freedom, and many different computational and mechanical and sensory subsystems another problem arises. Whereas with small behavior-based robots it was rather clear what the robot had to do at any point, and very little chance for conflict between subsystems, this is not the case with a humanoid robot.

Suppose the humanoid robot is trying to carry out some manipulation task and is foveating on its hand and the object with which it is interacting. But, then suppose some object moves in its peripheral vision. Should it saccade to the motion to determine what it is? Under some circumstances this would be the appropriate behavior, for instance when the humanoid is just fooling around and is not highly motivated by the task at hand. But when it is engaged in active play with a person, and there is a lot of background activity going on in the room this would be entirely inappropriate. If it kept saccading to everything moving in the room it would not be able to engage the person sufficiently, who no doubt would find the robot's behavior distracting and even annoying.

This is just one simple example of the problem of coherence. A humanoid robot has many different subsystems, and many different low level reflexes and behavioral patterns. How all these should be orchestrated, especially without a centralized controller, into some sort of coherent behavior will become a central problem.

2.4 Self adaptation

When we want a humanoid robot to act fluently in the world interacting with different objects and people it is a very different situation to that in classical robotics (Brooks 1991) where the robot essentially goes through only a limited number of stereotyped actions. Now the robot must be able to adapt its motor control to changes in the dynamics of its interaction with the world, variously due to changes in what it is grasping, changes in relative internal temperatures of its many parts — brought about by the widely different activities it engages in over time, drift in its many and various sensors, changes in lighting conditions during the day as the sun moves, etc. Given the wide variety of sensory and motor patterns expected of the robot it is simply not practical to think of having a fully calibrated system where calibration is separate from interaction. Instead the system must be continuously self adapting, and thus self calibrating. The challenge is to identify the appropriate signals that can be extracted from the environment in order to have this adaptation happen seamlessly behind the scenes.

Two classical types of learning that have been little used in robotics are habituation and sensitization. Both of these types of learning seem to be critical in adapting a complex robot to a complex environment. Both are likely to turn out to be critical for self adaptation of complex systems.

2.5 Development

The designers of the hardware and software for a humanoid robot can play the role of evolution, trying to instill in the robot the resources with which evolution endows human individuals. But a human baby is not a developed human. It goes through a long series of developmental stages using scaffolding (see (Rutkowska 1994) for a review) built at previous stages, and expectation-based drives from its primary care givers to develop into an adult human. While *in principle* it might be possible to build an adult-level intelligence fully formed, another approach is to build the baby-like levels, and then recapitulate human development in order to gain adult human-like understanding of the world and self. Such coupling of the development of cognitive, sensory and motor developments has also been suggested by Pfeiffer (1996). Cognitive development is a completely new challenge for robotics, behavior-based or otherwise. Apart from the complexities of building development systems, there is also the added complication that in humans (and, indeed, most animals) there is a parallel development between cognitive activities, sensory capabilities, and motor abilities. The latter two are usually fully formed in any robot that is built so additional care must be taken in order to gain the natural

advantage that such lockstep development naturally provides.

2.6 Historical contingencies

In trying to understand the human system enough to build something which has human like behavior one always has to be conscious of what is essential and what is accidental in the human system. Slavishly incorporating everything that exists in the human system may well be a waste of time if the particular thing is merely a historical contingency of the evolutionary process, and no longer plays any significant role in the operation of people. On the other hand there may be aspects of humans that play no visible role in the fully developed system, but which act as part of the scaffolding that is crucial during development.

2.7 Inspiration from the brain

In building a human-level intelligence, a natural approach is to try to understand all the constraints that we can from what is known about the organization of the human brain. The truth is that the picture is entirely fuzzy and undergoes almost weekly revision as any scanning of the weekly journals indicates. What once appeared as matching anatomical and functional divisions of the brain are quickly disappearing, as we find that many different anatomical features are activated in many different sorts of tasks. The brain, not surprisingly, does not have the modularity that either a carefully designed system might have, nor that a philosophically pure, wished for, brain might have. Thus skepticism should be applied towards approaches that try to get functional decompositions from what is known about the brain, and trying to apply that directly to the design of the processing system for a humanoid robot.

3 The Humanoid Cog

Work has proceeded on Cog since 1993.

3.1 The Cog robot

The Cog robot, figure 1, is a human sized and shaped torso from the waist up, with an arm (soon to be two), an onboard claw for a hand (a full hand has been built but not integrated), a neck, and a head.

The torso is mounted on a fixed base with three degree of freedom hips. The neck also has three degrees of freedom. The arm(s) has six degrees of freedom, and the eyes each have two degrees of freedom.

All motors in the system are controlled by individual servo processors which update a PWM chip at 1000Hz. The motors have temperature sensors mounted on them, as do the driver chips — these, along with current sensors, are available to give a kinesthetic sense of the body and its activities in the world. Those motors on the eyes, neck, and torso all drive joints which have limit switches attached to them. On power up, the robot subsystems must drive the joints to their limits to calibrate the 16 bit shaft encoders which give position feedback.

The eyes have undergone a number of revisions, and will continue to do so. They form part of a high performance active vision system. In the first two versions each eye had a separate pan and tilt motor. In the third version a single tilt motor is shared by both eyes. Much care has been taken to reduce the drag from the video cables to the cameras. The eyes are capable of performing saccades with human level speeds, and similar levels of stability. A vestibular system using three orthogonal solid state gyroscopes and two inclinometers in now mounted within the head. The eyes themselves each consist of two black and white cameras mounted vertically one above the other.

The lower one is a wide angle peripheral camera (120' field of view), and the upper is a foveal camera (23' field of view).

The arm demonstrates a novel design in that all the joints use series elastic actuators (Williamson 1995). There is a physical spring in series between each motor and the link that it drives. Strain gauges are mounted on the spring to measure the twist induced in it. A 16 bit shaft encoder is mounted on the motor itself. The servo loops turn the physical spring into a virtual spring with a virtual spring constant. As far as control is concerned one can best think of controlling each link with two opposing springs as muscles. By setting the virtual spring constants one can set different equilibrium points for the link, with controllable stiffness (by making both springs proportionately more or less stiff).

The currently mounted hand is a touch sensitive claw.

The main processing system for Cog is a network of Motorola 68332s. These run a multithreaded lisp, L (written by the author), a downwardly compatible subset of Common Lisp. Each processing board has eight communications ports. Communications can be connected to each other via flat cables and dual ported RAMs. Video input and output is

connected in the same manner from frame grabbers and to display boards. The motor driver boards communicate via a different serial mechanism to individual 68332 boards. All the 68332's communicate with a Macintosh computer which acts as a file server and window manager.

3.2 Experiments with Cog

Engineering a full humanoid robot is a major task and has taken a number of years. Recently, however, we have started to be able to make interesting demonstrations on the integrated hardware. There have been a number of Master theses, namely Irie (1995), Marjanovic (1995), Matsuoka (1995a), and Williamson (1995), and a number of more recent papers Matsuoka (1995b), Ferrell (1996), Marjanovic, Scassellatti & Williamson (1996), and Williamson (1996), describing various aspects of the demonstrations with Cog.

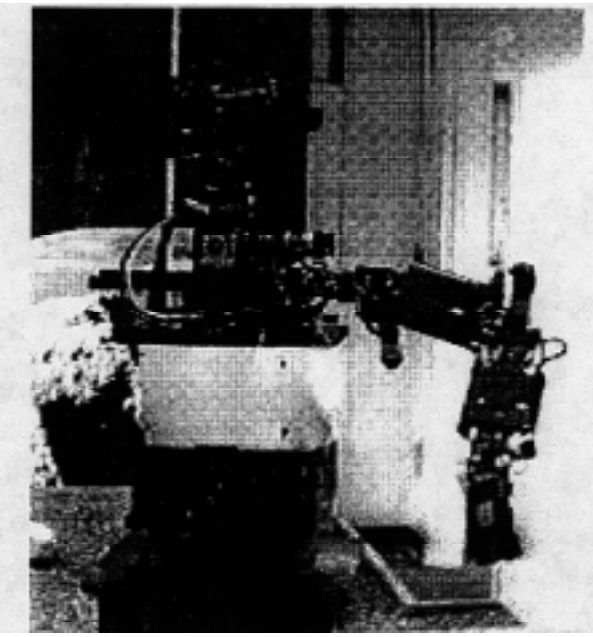

Figure 1: The robot Cog, looking at its left arm stretched out to one of its postural primitive positions.

Irie has built a sound localization system that uses cues very similar to those used by humans (phase difference below 1.5Khz, and time of arrival above that frequency) to determine the direction of sounds. He then used a set of neural networks to learn how to correlate particular sorts of sounds, and their apparent aural location with their visual location using the assumption that aural events are often accompanied by motion. This learning of the correlation between aural and visual sensing, along with a coupling into the occulomotor map is akin to the process that goes on in the superior colliculus.

Marjanovic started out with a simple model of the cerebellum that over a period of eight hours of training was able to compensate for the visual slip induced in the eyes by neck motions.

Matsuoka built a three fingered, one thumbed, self contained hand (not currently mounted on Cog). It is fully self contained and uses frictionally coupled tendons to automatically accomodate the shape of grasped objects. The hand is covered with touch sensors. A combination of three different sorts of neural networks is used to control the hand, although more work remains to be done to make then all incremental and on-line.

Williamson controls the robot arm using a biological model (from Bizzi, Giszter, Loeb, Mussa-Ivaldi & Saltiel (1995) based on postural primitives. The arm is completely compliant and is safe, for people to interact with. People can push the arm out of the way, just as they could with a child if they chose to.

The more recent work with Cog has concentrated on component behaviors that will be necessary for Cog to orient, using sound localization to a noisy and moving object, and then bat at the visually localized object. Ferrell has developed two dimensional topographic map structures which let Cog learn mappings between peripheral and foveal image coordinates and relate them to occulomotor coordinates. Marjanovic, Scassellati and Williamson, have used similar sorts of maps to relate hand coordinates to eye coordinates and can learn how to reach to a visual target. Their key, innovation is to cascade learned maps on top of each other, so that once a saccade map is learned, a forward model for reaching is learned using the saccade map to imagine where the eyes would have had to look in a zero error case, in order to deduce an observed error. The forward model is then used (all on-line and continuously, of course) to a train a backward model, or ballistic map, using the same imagination trick, and this then lets the robot reach to any visual target.

These basic capabilities will form the basis for higher level learning and development of Cog.

Acknowledgements

The bulk of the engineering on Cog has been carried out by graduate students Cynthia Ferrell, Matthew Williamson, Matthew

Maijanovic, Brian Scassellati, Robert Irie, and Yoky Matsuoka. Earlier Michael Binnard contributed to the body and neck, and Eleni Kappogiannis built the first version of the brain. They have all labored long and hard. Various pieces of software have been contributed by Mike Wessler, Lynn Stein, and Joanna Bryson. A number of undergraduates have been involved, including Elmer Lee, Erin Panttaja, Jonah Peskin, and Milton Wong. James McLurkin and Nicholas Schectman have contributed engineering assistance to the project. Rene Schaad recently built a vestibular system for Cog.

References

Bizzi, E., Giszter. S. F., Loeb, E., Mussa-Ivaldi, F. A. & Saltiel, P. (1995), "Modular Organization of Motor Behavior in the Frog's Spinal Cord", *Trends in Neurosciences* 18, 442-446.

Brooks, R. A. (1991), "New Approaches to Robotics", *Science* 253, 1227-1232.

Ferrell, C. (1996), "Orientation Behavior Using Registered Topographic Maps", in *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts. To appear.

Irie, R. E. (1995), "Robust Sound Localization: An Application of an Auditory Perception System for a Humanoid Robot", Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Johnson, M. (1987), *The Body in the Mind & The Bodily Basis of Meaning, Imagination, and Reason*, The University of Chicago Press, Chicago, Illinois.

Lakoff, G. (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago, Illinois.

Marjanovic, M. J. (1995), "Learning Maps Between Sensorimotor Systems on a Humanoid Robot", Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Marjanovic, M., Scassellatti, B. & Williamson, M. (1996), "Self-Taught Visually-Guided Pointing for a Humanoid Robot", in *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts. To appear.

Matsuoka, Y. (1995a), "Embodiment and Manipulation Learning Process for a Humanoid Hand", Technical Report TR-1546, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Matsuoka, Y. (1995b), "Primitive Manipulation Learning with Connectionism", in M.C.D.S. Touretzky & M. Hasselmo, eds, *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge, Massachusetts, pp. 889-895.

Pfeiffer, R. (1996), Building "'Fungus Eaters': Design Principles of Autonomous Agents", in *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts. To appear.

Rutkowska, J. C. (1994), "Scaling Up Sensorimotor Systems: Constraints from Human Infancy", *Adaptive Behavior 2*(4), 349-373.

Williamson, M. (1996), "Postural Primitives: Interactive Behavior for a Humanoid Robot Arm", in *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts. To appear.

Williamson, M. M. (1995), "Series Elastic Actuators", Technical Report 1524, Massachusetts Institute of Technology, Cambridge, Massachusetts.