Natural Tasking of Robots Based on Human Interaction Cues

PI: Rodney A. Brooks
1999–2004
DARPA MARS Program, BAA 99-09

A full website can be found at `http://www.ai.mit.edu/projects/mars/`.

All papers, these, and presentations referenceed in the report can be found in full in the subdirectory with this report, `papers`, as file `refxx.pdf`, where `xx` is the two digit reference number.

1 Overview

When a person gives a task to another person there are at least two sorts of very human processes at work. At the surface level, each person both displays and perceives cross-cultural cues which regulate the interaction. Through facial expressions, body posture, and utterances, the student unconsciously speeds or slows the rate at which the instructor is teaching and directs the instructor to provide more information when necessary. At a deeper level, the transfer of information is successful because both student and instructor share a common sense of how the world works. Both student and teacher share not only knowledge about how objects behave (an intuitive physics) but also knowledge about how other people behave (an intuitive psychology). Our challenge is to make commanding robots as intuitive and natural as commanding professional soldiers by providing a natural and intuitive interface that capitalizes on a person's intuitive understanding of how to communicate, and by instilling into robots that same deep understanding of the world which is shared by people.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JUN 2005	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Natural Tasking of Robots Based on Human Interaction Cues	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **The Massachusetts Inst Of Tech Cambridge Lab For Computer Science**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001779., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**28**	

2 Approach

We proposed developing the perceptual and intellectual abilities of robots so that in the field, war-fighters can interact with them in the same natural ways as they do with their human cohorts. To illustrate our goals and illuminate the technical problems that we must solve to achieve these goals, we will outlined three scenarios:

1. Showing a robot how to open the gas tank of an unfamiliar captured enemy vehicle.

2. Instructing a robot to carry out a reconnaissance mission ranging a few hundred meters from the command post within a strife-torn downtown urban environment.

3. Instructing a dextrous forklift-like robot to load a truck by showing it how the particular bulk food sacks should be stacked together, one by one.

Our approach was based on two key ideas; imitation and social interaction.

Imitation involves the robot watching and listening to a person perform some task and then equivalently executing it. From its observations the robot must extract which aspects of the person's motions and utterances are essential to actually carrying out the task, which are part of the instruction but not part of the actual task, and which are simply connective or coincidental.

Social interaction involves the robot engaging a person in the same dynamic two-way communication processes which two people could share. Each participant gives the other subconscious cues that carry messages such as "I understand that", "you're going too fast", "I don't know what you mean", "I already know that", "look at what's important", "no, it's more like this", "you don't understand it", and "now you get it!". The mechanisms for these signals are complex and often interrelated and involve such indications as gaze direction, eye contact, averting eye contact, nodding, body posture shifts, facial expressions, head motions, pre-linguistic verbalizations ("hmmm" or "uh-huh"), and codified verbalizations ("Sir!").

The principles of *development*, *embodimen* and *integration* contributed to our approach. The process of development wherein humans perform incrementally more difficult tasks in complex environments as they mature inspires a developmental methodology for our robots. Embodiment emphasizes human-like aspects of our robots' bodies. The integration of multiple sensory modalities, physical degrees of freedom and behavioral systems all a single robot to imitate and interact with humans in a more sophisticated manner.

3 Research Questions

In trying to use imitation and social interaction techniques for human-robot communication and for tasking robots in the field, there arise at least six deep and difficult questions, each of which has many technical components which form the topics on which we propose

1. Knowing what aspects of behavior to imitate.
2. Mapping from one body to another.
3. Implementing corrective actions and recognizing success.
4. Chaining pieces of action together into larger tasks.
5. Generalizing imitated actions to different and more complex tasks.
6. Making the interactions intuitive for the human.

The chart below displays examples of using the principles of social interaction, development, embodiment and system integration to address the six major research questions we have identified.

	Knowing what to imitate (1)	Mapping between bodies (2)	Correcting failures and recognizing success (3)	Chaining actions (4)	Generalizing to more complex tasks (5)	Intuitive Interactions (6)
Social Interaction (I)	Uses attentional cues to recognize task relevant events and objects		Identifies and displays emotional states for recognizing success			Allows humans and robots to share similar social cues without effort
Development (II)	Limits search space by incremental refinement of perception	Simplifies mapping by incremental refinement of motor skills		Provides natural decomposition of complex tasks	Exploits incremental learning to build new, more complex skills	Provides an established framework for building social skills
Embodiment (III)	Assists directed perception by constraining possible movements	Provides similar structure which simplifies mapping between bodies	Enables simple but robust low-level behaviors	Places physical limits on sequential actions		Allows the human to observe the robot's natural social cues
Integration (IV)	Allows robot to recognize cues in multiple modalities (voice, gesture)		Increases robustness through multi-channel redundancy		Assists in transfer to new modalities	Allows human to use natural methods of communication (voice, gesture)

4 Student Output

Students supported by this effort for the PhD degrees have gone on to a number of positions:

- Artur M. Arsenio, researcher, Siemens.
- Cynthia Breazeal, Assistant Professor of Media Arts and Sciences, MIT.
- Paul Fitzpatrick, Lecturer, Electrical Engineering and Computer Science, MIT.
- Charles C. Kemp, post-doctoral associate, CSAIL, MIT.
- Matthew Marjanovic, researcher, ITA Software.
- Brian Scasselatti, Assistant Professor of Computer Science, Yale.
- Matthew Williamson, research, Sana Security.

Postdoctoral students supported by this effort have moved on to new positions:

- Martin C. Martin, researcher, Icosystems.
- Giorgio Metta, Lecturer, University of Genoa.

A number of students who started work under this effort continue their PhD studies at MIT:

- Bryan Adams
- Lijin Aryananda
- Jessica Banks
- Aaron Edsinger-Gonzales
- Eduardo Torres-Jara
- Paulina Varchavskaia

5 Results for 1999–2000

Flexible Turn-Taking based on eye contact and head motion. We have demonstrated robust and flexible vocal turn-taking on our robot, Kismet. Kismet can engage in a proto-dialog with a single person as well as with two people. Kismet determines when it should take its turn based on pauses in speech and the current phase of the turn-taking interaction. Through experiments with naive subjects, we have found that people intuitively read the robot's physical and vocal cues (change of gaze direction, shifts of posture, and pauses in vocalizations) and naturally use these cues to time their own response. As a result, the proto-dialog becomes smoother over time, with fewer accidental interruptions or pauses.

Detect Prosody in human speech and show appropriate facial responses. We have demonstrated a robust technique for recognizing affective intent in robot-directed speech. By analyzing the prosody of a person's speech, Kismet can determine whether it was praised, prohibited, soothed, or given an attentional bid. The robot can distinguish these affective intents from neutral robot-directed speech. The output of the recognizer modulates the robot's emotional models, inducing an appropriate affective state with

a corresponding facial expression (an expression of happiness when praised, sorrow when prohibited, interest when alerted, and a relaxed expression for soothing). In multi-lingual experiments with naive female subjects, we found that the robot was able to robustly classify the four affective intents. In addition, the subjects intuitively inferred when their intent had been properly understood by Kismet's expressive feedback.

Expressive feedback through face, voice, and body posture. We have implemented expressive feedback in multiple modalities on Kismet. The robot is able to express itself through voice, facial expression, and body posture. We have evaluated the readability of Kismet's expressions for anger, disgust, fear, happiness, interest, sorrow, surprise, and some interesting blends through numerous studies with naive human subjects.

Visual attention and gaze direction. We have implemented a visual attention system on Cog and Kismet based on Jeremy Wolfe's model of human visual search. We have tested the robustness of the attention system on these robots. By matching the robot's visual system to what humans find to be inherently salient, the robot's attention is often drawn to the same sorts of stimuli that humans do. In studies with naive subjects, we found that people intuitively use natural attention-grabbing cues to quickly direct the robot's attention (motion, proximity, etc.). The subject's intuitively use the robot's gaze and smooth pursuit behavior to determine when they have successfully directed the robot's attention.

Papers on this work included: [1], [47], [4], [35], [30], [37], [38], [85], [86], [48], [84], [39], [40], [98].

Presentations on this work included: [43], [31], [42].

6 Results for 2000-2001

Detecting Head Orientation. We have implemented and evaluated a system that detects the orientation of a person's head from as far as six meters away from the robot. To accomplish this, we have implemented a multi-stage behavior. Whenever the robot sees an item of interest, it moves its eyes and head to bring that object within the field of view of the foveal cameras. A face finding algorithm based on skin color and shape is used to identify faces and a software zoom is used to capture as much information

as possible. The system then identifies a set of facial features (eyes and nose/mouth) and uses a model of human facial structure to identify the orientation of the person's head.

Mimicry. Cog's torso was retrofitted with force sensing capabilities in order to implement body motion via virtual spring force control. In addition, we developed a representational language for humanoid motor control inspired by the neurophysiological organizing principle of motor primitives. Both endeavors allowed Cog to broadly mimic the motions of a person with whom it interacts using its body or arms. In the arm imitation behavior, the robot continuously tracks many object trajectories. A trajectory is selected on the basis of animacy and the attentional state of the instructor. Motion trajectories are then converted from a visual representation to a motor representation which the robot can execute.The performance of this mimicry response was evaluated with naive human instructors.

Distinguishing Animate from Inanimate. We have implemented a system that distinguishes between the movement patterns of animate objects from those of inanimate objects. This system uses a multi-agent architecture to represent a set of naive rules of physics that are drawn from experimental results on human subjects. These naive rules represent the effects of gravity, inertia, and other intuitive parts of Newtonian mechanics. We have evaluated this system by comparing the results to human performance on classifying the movement of point-light sources, and found the system to be more than 85% accurate on a test suite of recorded real-world data.

Joint Reference. Using its new 2-DOF hands that exploit series elastic actuators and rapid prototyping technology, Cog demonstrated basic grasping and gestures. The gestural ability was combined with models from human development for establishing joint reference, that is, for the robot to attend to the same object that an instructor is attending to. Objects that are within the approximate attention range of the human instructor are made more salient to the robot. Information from head orientation is the primary cue of attention in the instructor.

Simulated Musculature. Cog's arm and body are controlled via simulated muscle-like elements that span multiple joints and operate indepen-

dently. Muscle strength and fatigue over time are modulated by a biochemical model. The muscle-like elements are inspired by real physiology and allow Cog to move with dynamics that are more human-like than conventional manipulator control.

Vocabulary Management. Kismet needs to acquire a vocabulary relevant to a human's purpose. Towards this goal, first, we have implemented a command protocol for introducing vocabulary to Kismet. Second, we have developed an unsupervised mechanism for extracting candidate vocabulary items from natural continuous speech. Third, we have analyzed the speech used in teaching Kismet words in order to determine whether humans naturally modify their speech in ways that would enable better word learning by the robot.

Head Pose Estimation. We developed a fully automatic system for recovering the rigid components of head pose. The conventional approach of tracking pose changes relative to a reference configuration can give high accuracy but is subject to drift. In face-to-face interaction with a robot, there are likely to be frequent presentations of the head in a close to frontal orientation, so we used that to make opportunistic corrections. Tracking of pose was done in an intermediate mixed coordinate system chosen to minimize the impact of errors in estimates of the 3D shape of the head being tracked. This is vital for practical application to unknown users in cluttered conditions.

Papers on this work included: [2], [36], [32], [33], [34], [41], [53], [51], [78], [89], [88], [87], [96].

Presentations on this work included:
[3], [44], [52], [75], [79], [90], [91], [94].

7 Results for 2001-2002

Cog

Adaptation of Arm Stiffness. Cog learns a feed forward command force function that is dependent on arm posture but independent of stiffness. This adaptation of stiffness parallels human reaching in which there is higher stiffness at the endpoints and lower stiffness during the middle of a reach. It allows Cog to reach to points in the arm's workspace with greater accuracy,

gives Cog a more human-like range of dynamics and allows for safer and more intuitive physical interaction with humans.

Reflex Inhibition. Inhibition of extreme movements prevents robotic failure. Cog uses learned reflex inhibition for coordinated joint movement and distribution of a movement over as many degrees of freedom as possible, avoiding saturation of a few joints. During learning Cog explores the gross limits of its torso workspace by the action of reflexive movements. As it reaches joint extremities, a simulated pain model results in modification of a reflex to constrain its movements to avoid physically harming itself and to operate the torso primarily in a state of balance.

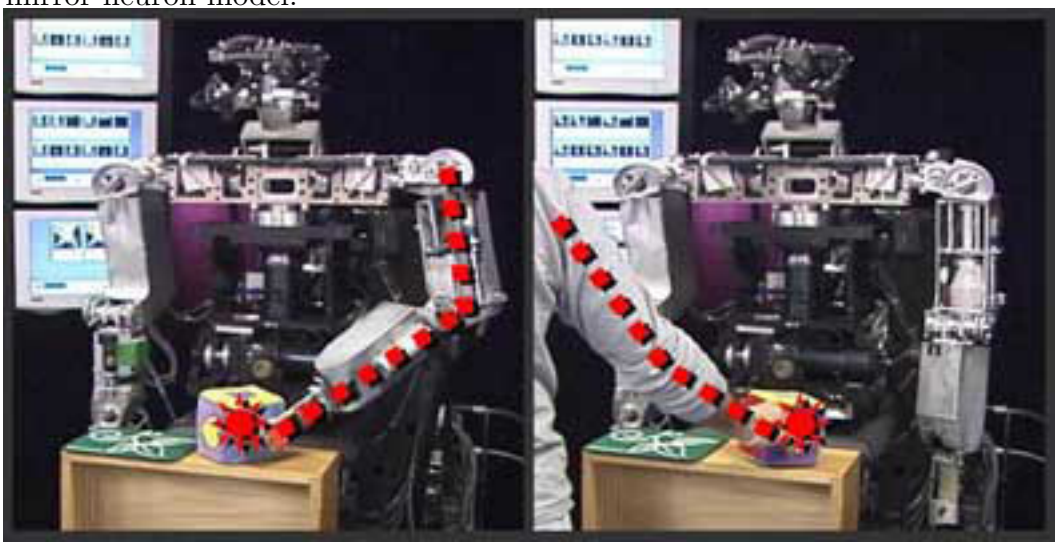
Dynamic Configuration of Multi-joint Muscles. To facilitate development of a multi-joint muscle model for controlling Cog, a graphical user interface (GUI) displays the movement of Cog in terms of Cog's muscle model overlaying Cog's joints. The muscle model is reconfigurable at run time through the GUI.

Hand Reflex. Cog's two degrees of freedom hand, equipped with tactile sensors, has a reflex that grasps and extends in a manner similar to primate infants. Contact inside the hand causes a short term grasp, contact to the back of the hand causes an extensive stretch.

Arm Localization. It is difficult to visually distinguish the motion of a robot's own arm as distinct from similar motion by humans or objects. Cog discovers and learns about its own arm by generating a motion and then correlating the associated optic flow with proprioceptive feedback. It ignores any uncorrelated movements and visual data. Once Cog can track its own arm, when it contacts an object, it discounts its own movement in order to isolate object properties.

Object Tapping for Segmentation. There are cases when solely visual based object segmentation poorly or completely fails to disambiguate an object from its background. Cog can determine the shape of simple objects by tapping them. This physical experimentation augments visual based segmentation.

Mirror-Neuron Model. Cog is able to perform manipulative actions: poking an object away from its body and poking an object towards itself. It uses its attentional system to locate and fixate an object and its tracking system to follow the object trajectory. It maps visual perception into a sequence of motor commands to engage the object. These abilities: vision driven manipulation and mapping perception to action are prerequisites of a mirror neuron model.



On the left, the robot establishes a causal connection between commanded motion and its own manipulator, and then probes its manipulator's effect on an object. The object then serves as a literal "point of contact" to link robot manipulation with human manipulation (on the right), as is required for a mirror-neuron-like representation.

Module Integration. Cog has a modular architecture with components responsible for sensing, acting and processing higher level aspects of vision and manipulation. Cog integrates modules responsible for 14 degrees of freedom (head, torso and arm axes) in order to reach out and poke an object. It coordinates its head control and arm control with its visual attention, tracking, and arm localization subsystems.

Face Tracking. Cog's attentional system is updated with an imported face detector that has greater accuracy. The detector is coupled with a face tracker that copes with non-frontal face presentations despite the detector

operating slower than frame rate. The combined systems allow Cog to engage in tasks requiring shared attention and human-robot interaction.

M4

7.1 Macaco.

The M4 robot consists of an active vision robotic head integrated with a Magellan mobile platform. The robot integrates vision-based navigation with human-robot interaction. It operates a portable version of the attentional systems of Cog and Lazlo with specific customization for a thermal camera. Navigation, social preferences and protection of self are fulfilled with a model of motivational drives. Multi-tasking behaviors such as night time object detection, thermal-based navigation, heat detection, obstacle detection and object reconstruction are based upon a competition model.

Kismet

Dynamic Subjective Response. Kismet has the ability to learn to recognize and remember people it interacts with. Such social competence leads to complex social behavior, such as cooperation, dislike or loyalty. Kismet has an online and unsupervised face recognition system, where the robot opportunistically collects, labels, and learns various faces while interacting with people, starting from an empty database.

Proto-linguisitc Capabilities. Kismet uses utterances as a way to manipulate its environment through the beliefs and actions of others. It has a vocal behavior system forming a pragmatic basis for higher level language acquisition. Protoverbal behaviors are influenced by the robot's current perceptual, behavioral and emotional states. Novel words (or concepts) are created and managed. The vocal label for a concept is acquired and updated.

Papers on this work included:
[28], [68], [55], [81], [83], [97], [95].
Presentations on this work included:
[6], [45], [54], [80].

8 Results for 2002-2003

Guided Training via a Modular Software System for Learning from Interaction with the Environment and People. Cog learns simple arm and end effector tasks via a combination of self-exploration and explicit training. With tactile reinforcement signals, Cog is taught by a human trainer to perform simple postural arm and hand actions. Subsequently, the trainer teaches the robot to perform such learned actions in response to tactile (touch to particular fingers) and visual (objects of particular colors) stimuli.

Expoiting a Model of Muscle Fatique for Human-like Movement. Cog has a fatigue model for its virtual musculature. This simulation of biological muscle fatigue provided signals that modulated motor performance and provided negative reinforcement to the learning module to guide the acquisition of more natural human-like motor movement.

Learning How Joints Move in Relation to Virtual Muscle Groups. Starting simply, from an inclination to randomly move its virtual muscles, Cog learns to activate its muscle model so it can move to particular points in joint angle space. Cog acquires an unsupervised linear dependency model between joint velocities and controller modules that supervise multiple muscles in combination.

Active segmentation. Cog uses active exploration to resolve visual ambiguity in its workspace. Objects can sometimes be difficult to locate if their visual appearance is similar to the general background. Cog solves this problem by sweeping its arm through regions of interest. If no object is there, the arm passes unimpeded. If an object is present, the impact between it and the robot's arm causes the object to move, revealing its boundary.

Cog uses a mirror neuron model to learn how different objects respond to the actions it can perform. If the robot taps an object and it slips and rolls, it learns to predict the direction of slip based on visaul evidence, and can then use that information to deliberately trigger or avoid rolling an object while tapping it. The mirror neuron model allows the robot to mimic an action demonstrated by a human relative to the natural behavior of the object, rather than pure geometry.

Open object recognition. With open object recognition, the set of objects Cog can recognize grows over time, as it accumulates experience through active segmentation and other experimental methods. The robot clusters episodes of object interaction to learn the properties of novel, unfamiliar objects. An operator can introduce names for objects to facilitate further task-related communication.

Perceptual cycle. Cog uses the constraints of known activities to learn about the objects used within those activities – for example, during manipulation. Cog can track known objects to learn about activities they occur in, such as a sorting task or object search. By combining the ability to learn about objects through activity constraints and activities through tracking objects, the robot can achieve a virtuous cycle of perception.

Adaptive control of Cog's arm using a nonlinear sliding-modes controller. Two degrees of freedom on Cog's arm operate via non-parametric adaptive control using a nonlinear sliding-modes controller. This sufficiently mitigates the high signal to noise ratio arising in Cog's arm (due to a small strain gauge signal that experiences capacitive coupling with other signals) and allows semi-autonomous, task adequate control.

Learning actions and objects from observed use. While Cog watches an event involving someone's arm handling an object (e.g. filing a surface, swinging a pendulum), its vision system extracts both the nature of the arm movement and derives a predictive dynamical model of the object.

A compact linear series elastic actuator design for human-like neck joint. For a new robotic head, two new coupled neck axes were designed and built using linear series elastic actuators aligned in parallel. The design is compact: the two axes have intersecting centers of rotation. Force control in combination with elastic actuation provides safe, human like compliancy.

The ALIVE architecture. The ALIVE architecture consisting of a stack and the CreaL software development environment controls the new robotic head. The stack is a special purpose, extensible, real-time, small form-factor hardware architecture of controller boards, sensor boards, network board, and off-the-shelf processor. CreaL, which is retargetable, extracts efficient

computational power to allow many lightweight threads from the relatively cheap off-the-shelf processor via efficient software scheduling, compilation and language abstraction. The ALIVE architecture facilitates complete designer control over startup and failure sequences which is essential for continuous, safe robot operation.

Papers on this work included: [8], [9], [11], [12], [10], [7], [57], [58], [69], [60], [62], [63], [67], [70], [72], [76], [77], [82].

Presentations on this work included:
[5], [56], [61], [64], [59].

9 Results for 2003-2004

Accomplishments are on 7 robotic platforms: Cog, Cardea, Coco, a robot head named Mertz, a new humanoid named Domo, a human wearable/hybrid system named Duo and an unnamed 5 DOF hand. This work was done between July 2003 and July 2004.

The 'Yet Another Robot Platform' open software library is used on 6 platforms. Software written in C and C++ that provides routines for robot platform development in terms of inter-process communication, vision and control and has operating system services support for Windows NT and QNX4 and QNX6 is running on multiple robotic platforms at MIT: Cog, Coco, Domo, Mertz, Cardea and Duo, and in Europe.

Door Shoving by A Self-Balancing Mobile Humanoid. Cardea , a prototype humanoid based on a Segway RMP extended with contact and IR sensing, simple torso and 3 DOF arm manipulation, vision using 1 fixed mounted camera and a single DOF camera and entirely on-board computation, can navigate an office corridor, find a partially closed door, shove it open and pass through.

Emergency Kickstands within Safety System for Segway RMP base. Two emergency kickstands for Cardea deploy when a 'sniffer' detects software definable error conditions indicating the platform is falling over. They are part of a complete safety system that overrides robotic control when the RMP over-tilts. First, the system relies on RMP self-balancing. When

self-balancing fails, the kickstands eject. Safety is also ensured via radio controlled Emergency stop (E-stop).

A Lightweight Computational Hardware Architecture Supporting Humanoid Mobility and Manipulation. A computational hardware architecture consisting of a network of distributed, onboard lightweight 8-bit computational elements that supports behavior, sensorimotor and RMP controllers, power circuitry and debugging demonstrates humanoid navigation and manipulation.

A Prototype Camera-Arm Platform Integrating a Visual System and a Motor System Running on an Embedded Architecture The design of embedded brushless motor amplifiers, DSP motor controllers and sensor conditioning is integrated with the ALIVE hardware and software architecture. A 5 DOF force controllable prototype arm, with series elastic actuators, a differentially driven shoulder and a virtually centered elbow, runs on the embedded architecture using virtual spring control and a 'CreaL' (creature language) behavioral controller. It can track in conjunction with a 2 DOF active vision system running on a laptop. It can reach towards and poke an object using visual and color information and estimating the position of its hand via forward kinematics in visual coordinate space.

A Creature-based Approach to Robotic Existence. Mertz, an active-vision humanoid head platform, fulfills an immediate goal of running continuously for days without supervision at a variety of locations. Mertz is designed with fault prevention strategies in mind, It can instantly startup and perform joint calibration. It has circuitry to protect against power cycles and abrupt shutdown. Its vision system is adaptable to different lighting conditions and backgrounds.

Domo: A Force Sensing and Compliant Humanoid Platform. Completed the design, fabrication and assembly of a new force sensing and compliant humanoid platform, named Domo, for exploring general dextrous manipulation, visual perception and learning. Domo incorporates force sensors and compliance in most of its joints to act safely in an unstructured environment. It consists of a two 6 DOF arms, two 4 DOF hands, a 7 DOF head, a 2 DOF neck, 58 proprioceptive sensors and 24 tactile sensors. Twenty-four

DOF use force controlled compliant actuators. Its realtime sensorimotor system is managed by an embedded network of DSP controllers. Its vision system, which (2 cameras, 3 DOF) utilizes the YARP software library, and its cognitive system run on a small, networked cluster of PC's.

Overcoming Mechanical Modes of Failure. Domo achieves mechanical robustness: geartrain failures are mitigated by using ball screws and elastic spring elements, motor winding reheating is avoided by current limits in its brushless DC motor amplifiers and prevention of stall currents, cable breakage and wire strain susceptibility have been reduced, and maintenance is easier by the design of modular subsystems.

Two Force Controlled Arms. Domo's arm design focuses on force control. An arm is passively or actively compliant and able to directly sense and command torques at each joint. This design forgoes the conventional emphasis on end effector stiffness and precision to, instead, mimic human capabilities. It relies on advanced linear Series Elastic Actuators.

A Robust Multi-Layered Sensorimotor and Cognitive System. Domo has been designed with four layers of sensorimotor and cognitive systems: physical for sensors, motors and interface electronics, DSP for real time control, a sensorimotor abstraction layer for interfacing between the DSP and cognitive layers, and a cognitive layer. It emphasizes robustness to common modes of failure, real-time control of time critical resources and expandable computational capability. This runs on a combination of special purpose embedded hardware communicating through a CAN bus or Firewire, in the case of cameras, to a cluster of Linux nodes.

Advanced Design of Elastic Force Sensing Actuators with Embedded Amplifiers. Design of new version of SEA using a) linear ball screws for greater efficiency and shock tolerance b) a cable drive transmission allowing actuator mass to be moved far from the end point reducing energy consumption and hence needing lower wattage motors, plus allowing modular and standardized packaging implying easier maintenance and reuse. A novel force sensing compliant (FSC) actuator places the spring element between the motor housing and the chassis ground which allows continuous rotation at the motor output. The FSC actuator is compact due to use of torsion

springs. Embedded custom brushless motor amplifiers and sensory signal amplifiers that reduce wiring run-length and thus simplify cable routing and lead to better robustness are incorporated.

A 5 DOF Sensor Rich hand with Series Elastic Actuation. Design, fabrication and assembly a 5 DOF sensor rich hand with simple, scalable force actuators . Three fingers with 8 force sensing axes and 5 position sensors , each consisting of 2 coupled and decoupling links driven by a compact, inexpensive rotary series elastic actuator which makes the hand mechanically compliant and force controllable. The last two links of each finger are equipped with dense arrays of force sensing resistors.

DUO: A Human/Wearable Hybrid for Learning About Common Manipulable Objects. Duo consists of a glasses mounted digital camera connected to a backpack holding a laptop which communicates wirelessly to a computer cluster. It also has four orientation sensors that are head, wrist, upper arm and torso mounted. Duo passively and actively observes the manipulation of objects in natural, unconstrained environments. It measures the kinematic configuration of its wearer's head, torso and dominant arm while watching its wearer's workspace through a head mounted camera. It requests helpful actions from its wearer through speech via headphones. It can segment common manipulable objects with high quality.

Using Cast Shadows for Visually-Guided Touching. The shadow cast by a robot's own body is used to help direct its arm towards, across, and away from an unmodeled surface without damaging it. The shadow is detected by a camera and used to derive a time-to-contact estimate which, when combined with the 2D tracked location of the arm's endpoint in the camera image is sufficient to allow 3D control relative to the surface.

Exploiting Amodal Cues for Robot Perception. Rhythmically moving objects, such as tools and toys, are detected, segmented and recognized by the sounds they generate as they move. This method does not require accurate sound localization but can complement that information. It is selective and robust in the face of distracting motion and sounds. This perceptual tool is required for a robot to learn to use tools and toys through demonstration.

Object Segmentation by Demonstration. A human teaches Cog how to segment objects from arbitrarily complex non-static images by waving the object to introduce it. An algorithm detects the skin color of the human's arm, and tracks its motion. Then the object's compact cover is extracted using the periodic trajectory information.

Figure/Ground Segmentation from Human Cues. In order to infer large scale depth and build 3-dimensional maps, Cog exploits its human helper's arm as a reference measure while measuring the relative size of objects on a monocular image. It is also able to perform figure/ground segregation on typical heavy objects in a scene, such as furniture and perform 3D object and scene reconstruction. This argues for solving a visual problem not simply by controlling the perceptual system, but actively changing the environment through experimental manipulation.

A Learning Framework for a Humanoid Robot Inspired by Developmental Learning. For Cog to learn about its physical surroundings, a human helps Cog to correlate its own senses, to control and integrate situational cues from its surrounding world and to learn about out-of-reach objects and the different representations in which they might appear. The strategies for this learning are inspired by child development theory which defines a separation and individuation developmental phase.

On-line Parameter Tuning of Neural Oscillators. Cog employs neural oscillators in its arm that are capable of adapting to the dynamics of the arm's controlled system. After using a time-domain analysis to intuitively tune the parameters of neural oscillators, Cog plays a rhythmic musical instrument such as a drum or tamborine.

Learning Task Sequences from Scratch. Task sequencing requires recognizing an object, identifying it with some associated action then learning the sequence of events and objects that characterize the task. For example, a saw must be recognized and moved back and forth on the correct plane to complete the task of sawing. Cog can learn task sequences from human-robot interaction cues. A human teaches the robot new objects such as tools and toys and their functionality. The robot explores the world and extends its

knowledge of the objects' properties. It acquires recognition of multi-modal percepts by manipulating the tools and toys.

Papers on this work included: [13], [20], [15], [22], [21], [24], [25], [26], [19], [16], [23], [14], [17], [18], [50], [49], [46], [27], [29], [65], [66], [71], [73], [93],

Presentations on this work included: [92], [74].

References

[1] Bryan Adams. Meso: A virtual musculature for humanoid motor control. Master's thesis, Massachusetts Institute of Technology, Department of Electrical and Computer Science, Cambridge, Ma, 2000.

[2] Bryan Adams. Learning humanoid arm gestures, March 2001. Working Notes - AAAI Spring Symposium Series: Learning Grounded Representations.

[3] Bryan Adams. Learning humanoid arm gestures. Presentation for AAAI Spring Symposium Series: Learning Grounded Representations, March 2001.

[4] Bryan Adams, Cynthia Breazeal, Rodney Brooks, and Brian Scassellati. Humanoid robots: A new kind of tool. *IEEE Intelligent Systems*, 2000.

[5] Artur Arsenio. Exploiting cross-modal rhythm for robot perception of objects. Presentation CIRAS.

[6] Artur Arsenio. Macaco: Acts and mind in accord. MIT AI Lab, Living Breathing Robots Group, March 2002.

[7] Artur Arsenio, Paul Fitzpatrick, Charles C. Kemp, and Giorgio Metta. The whole world in your hand: Active and interactive segmentation. *Third International Workshop on Epigenetic Robotics*, 2003.

[8] Artur M. Arsenio. Active vision for sociable, mobile robots. *In Proceedings of the Second International Conference on Computational Intelligence Robotics and Autonomous Systems - Special session in Robots with a Vision*, 2003.

[9] Artur M. Arsenio. Embodied vision - perceiving objects from actions. *IEEE International Workshop on Human-Robot Interactive Communication*, 2003.

[10] Artur M. Arsenio. Object segmentation through human-robot interactions in the frequency domain. *SIBGRAPI*, 2003.

[11] Artur M. Arsenio. A robot in a box. *11th International Conference on Advanced Robotics (ICAR'03)*, 2003.

[12] Artur M. Arsenio. Towards pervasive robotics. *Accepted for publication to the International Joint Conference on Artificial Intelligence*, 2003.

[13] Artur M Arsenio. Children, humanoid robots and caregivers. In *Fourth International Workshop on Epigenetic Robotics*, 2004.

[14] Artur M Arsenio. *Cognitive-Developmental Learning for a Humanoid Robot: A Caregiver's Gift*. PhD thesis, MIT, Cambridge, Ma., 2004.

[15] Artur M Arsenio. Developmental learning on a humanoid robot. In *IEEE International Joint Conference on Neural Networks*, Budapest, 2004.

[16] Artur M Arsenio. An embodied approach to perceptual grouping. In *IEEE CVPR Workshop on Perceptual Organization in Computer Vision*, 2004.

[17] Artur M Arsenio. Exploiting amodal cues for robot perception. Submitted to special issue of International Journal of Humanoid Robotics, 2004.

[18] Artur M Arsenio. Exploiting cross-modal rhythm for robot perception of objects. In *2nd International Conference on Computational Intelligence, Robotics, and Autonomous Systems*, Singapore, 2004.

[19] Artur M Arsenio. Figure/ground segregation from human cues. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[20] Artur M Arsenio. Learning task sequences from scratch: Applications to the control of tools and toys by a humanoid robot. In *IEEE Conference on Control Applications*, 2004.

[21] Artur M Arsenio. Map building from human computer interactions. In *IEEE CVPR Workshop on Real-Time Vision for Human Computer Interaction*, 2004.

[22] Artur M Arsenio. Object recognition from multiple percepts. In *IEEE-RAS/RSJ International Conference on Humanoid Robots*, 2004.

[23] Artur M Arsenio. On stability and tuning of neural oscillators: Application to rhythmic control of a humanoid robot. In *International Joint Conference on Neural Networks*, 2004.

[24] Artur M Arsenio. Teaching a humanoid robot from books. In *International Symposium on Robotics*, 2004.

[25] Artur M Arsenio. Teaching humanoid robots like children: Explorations into the world of toys and learning activities. In *IEEE-RAS/RSJ International Conference on Humanoid Robots*, 2004.

[26] Artur M Arsenio. Towards an embodied and situated ai. In *International FLAIRS Conference*, 2004. Nominated for Best Paper Award.

[27] Artur M. Arsenio and Paul Fitzpatrick. Exploiting cross-modal rhythm for robot perception of objects. In *2nd International Conference on Computational Intelligence, Robotics, and Autonomous Systems*, Singapore, 2003.

[28] Lijin Aryananda. Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Lausanne, Switzerland, 2002.

[29] Lijin Aryananda and Jeff Weber. Mertz: A quest for a robust and scalable active vision humanoid head robot. In *IEEE-RAS/RSJ International Conference on Humanoid Robots*, 2004.

[30] Cynthia Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 2000.

[31] Cynthia Breazeal. Sociable machines: Expressive social exchange between humans and robots. PhD Thesis Defense, May 2000.

[32] Cynthia Breazeal. Affective interaction between humans and robots. *The Sixth European Conference on Artificial Life (ECAL01)*, 2001.

[33] Cynthia Breazeal. Emotive qualities in robot speech. In *Proceedings of the 2001 IEEERSJ International Conference on Intelligent Robots and Systems (IROS01)*, Maui, HI, 2001.

[34] Cynthia Breazeal. Regulation and entrainment in human-robot interaction. *The International Journal of Experimental Robotics*, 2001.

[35] Cynthia Breazeal and Lijin Aryananda. Recognition of affective communicative intent in robot-directed speech. In *IEEE-RAS International Conference on Humanoid Robots*, 2000.

[36] Cynthia Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati. Active vision for sociable robots. *Socially Intelligent Agents - The Human in the Loop, Special Issue IEEE Transactions on Man, Cybernetics, and Systems, Part A: Systems and Humans*, 31(5):443–453, September 2001.

[37] Cynthia Breazeal, Aaron Edsinger, Paul Fitzpatrick, and Brian Scassellati. Social constraints on animate vision. In *IEEE-RAS International Conference on Humanoid Robots*, 2000.

[38] Cynthia Breazeal, Aaron Edsinger, Paul Fitzpatrick, Brian Scassellati, and Paulina Varchavskaia. Social constraints on animate vision. *IEEE Intelligent Systems*, July-August 2000.

[39] Cynthia Breazeal and Brian Scassellati. A context-dependent attention system for a social robot. *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, pages 1146–1151, 1999.

[40] Cynthia Breazeal and Brian Scassellati. How to build robots that make friends and influence people. *IROS99*, 1999.

[41] Cynthia Breazeal and Brian Scassellati. *Imitation in Animals and Artifacts*, chapter Challenges in Building Robots that Imitate People. MIT Press, 2001.

[42] Rodney Brooks Breazeal, Cynthia and Brian Scassellati. Natural tasking of robots based on human interaction cues. MARS Workshop, January 2000.

[43] Rodney Brooks. Natural tasking of robots based on human interaction cues - mit ai lab. DARPA Mobile Autonomouse Robot Software BAA9909, May 2000.

[44] Rodney Brooks. Natural tasking of robots based on human cues. Presentation for DARPA Mobile Autonomous Robot Software '01 PI Meeting, March 2001.

[45] Rodney Brooks. Natural tasking of robots based on human cues. DARPA Mobile Autonomous Robot Software, February 2002.

[46] Rodney Brooks, Lijin Aryananda, Aaron Edsinger, Paul Fitzpatrick, Charles Kemp, Una-May O'Reilly, Eduardo Torres-Jara, Paulina Varshavskaya, and Jeff Weber. Sensing and manipulating built-for-human environments. *International Journal of Humanoid Robotics*, 2004.

[47] Aaron Edsinger. A gestural language for a humanoid robot. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Ma, 2000.

[48] Aaron Edsinger and Una-May O'Reilly. Designing a humanoid robot face to fulfill a social contract. In *Proceedings of the IEEE International Workshop on Robot-Human Interaction*, 2000.

[49] Aaron Edsinger-Gonzales. Design of a compliant and force sensing hand for a humanoid robot. In *Proceedings of the International Conference on Intelligent Manipulation and Grasping*, July 2004.

[50] Aaron Edsinger-Gonzales and Jeff Weber. Domo: A force sensing humanoid robot for manipulation research. In *Proceedings of the 2004 IEEE International Conference on Humanoid Robots*, Santa Monica, Los Angeles, CA, USA., 2004. IEEE Press.

[51] Paul Fitzpatrick. From word-spotting to oov modelling, 2001. Term Paper for MIT Course 6.345.

[52] Paul Fitzpatrick. Head pose estimatin without manual initialization. Presentation for MIT Course 6.892, 2001.

[53] Paul Fitzpatrick. Head pose estimation without manual initialization, 2001. Term Paper for MIT Course 6.892.

[54] Paul Fitzpatrick. Better vision through poking. MIT AI Lab, Humanoid Robotics Group, June 2002.

[55] Paul Fitzpatrick. Role transfer for robot tasking. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, PhD Thesis Proposal, 2002.

[56] Paul Fitzpatrick. First contact: an active vision approach to object segmentation. Presentation IROS, 2003.

[57] Paul Fitzpatrick. First contact: An active vision approach to segmentation. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2003.

[58] Paul Fitzpatrick. *From First Contact to Close Encounters: A Developmentally Deep Perceptual System for a Humanoid Robot*. PhD thesis, MIT, 2003.

[59] Paul Fitzpatrick. From first contact to close encounters: building a developmentally deep perceptual system for a humanoid robot, phd thesis defense. MIT AI Lab, May 2003.

[60] Paul Fitzpatrick. Object lesson: Discovering and learning to recognize objects. In *3rd International Conference on Humanoid Robots*, Karlsruhe, Germany, October 2003.

[61] Paul Fitzpatrick. Object lesson: Discovering and learning to recognize objects. Presentation 3rd International IEEE/RAS Conference on Humanoids Conference, October 2003).

[62] Paul Fitzpatrick. Open object recognition for humanoid robots. *SPIE Robotics and Machine Perception newsletter*, 12(2):9, September 2003.

[63] Paul Fitzpatrick. Perception and perspective in robotics. *25th Annual Meeting of the Cognitive Science Society*, 2003.

[64] Paul Fitzpatrick. The whole world in your hand: Active and interactive segmentation. Presentation EPIROB 2003, August 2003.

[65] Paul Fitzpatrick. The dayone project: How far can a robot develop in 24 hours? In *Fourth International Workshop on Epigenetic Robotics*, Genoa, Italy, August 2004.

[66] Paul Fitzpatrick and Artur Arsenio. Feel the beat: Using cross-modal rhythm to integrate perception of objects, others, and self. In *Fourth International Workshop on Epigenetic Robotics*, Genoa, Italy, August 2004.

[67] Paul Fitzpatrick and Charles Kemp. Shoes as a platform for vision. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, pages 231–234, White Plains, New York, October 2003.

[68] Paul Fitzpatrick and Giorgio Metta. Towards manipulation-driven vision. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Lausanne, Switzerland, 2002.

[69] Paul Fitzpatrick and Giorgio Metta. Grounding vision through experimental manipulation. *Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences*, 2003.

[70] Paul Fitzpatrick, Giorgio Metta, Lorenzo Natale, Sajit Rao, and Giulio Sandini. Learning about objects through action - initial steps towards artificial cognition. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, May 2003.

[71] Paul Fitzpatrick and Eduardo Torres-Jara. The power of the dark side: Using cast shadows for visually-guided reaching. In *IEEE-RAS/RSJ International Conference on Humanoid Robots*, 2004.

[72] Charles C. Kemp. Duo: A human/wearable hybrid for learning about common manipulable objects. In *Proceedings of the 3rd International IEEE/RAS Conference on Humanoid Robots*, Karlsruhe, Germany, October 2003.

[73] Charles C. Kemp. Duo: A wearable system for learning about everyday objects and actions. In *8th IEEE International Symposium on Wearable Computers*, 2004.

[74] Charlie C. Kemp. Shoes as a platform for vision. 7th IEEE International Symposium on Wearable Computers, 2004.

[75] Matthew Marjanovic. meso: Simulated muscles for a humanoid robot. Presentation for Humanoid Robotics Group, MIT AI Lab, August 2001.

[76] Matthew J. Marjanovic. *Teaching an Old Robot New Tricks: Learning Novel Tasks via Interaction with People and Things*. PhD thesis, MIT, June 2003.

[77] Martin C. Martin. The essential dynamics algorithm: Essential results. Artificial Intelligence Memo AIM-2003-0014, Massachusetts Institute of Technology, May 2003.

[78] Giorgio Metta. An attentional system for a humanoid robot exploiting space variant vistion. *IEEE-RAS International Conference on Humanoid Robots 2001*, Nov 2001.

[79] Giorgio Metta. Lazlo's stuff. Presentatation for Living Breathing Robots Group, MIT AI Lab, August 2001.

[80] Giorgio Metta. Better vision through manipulation. Neuro-Engineering Workshop and Advanced School, June 2002.

[81] Giorgio Metta and Paul Fitzpatrick. Better vision through manipulation. In *Second International Workshop on Epigenetic Robotic*, Edinburgh, UK, August 2002.

[82] Giorgio Metta and Paul Fitzpatrick. Better vision through manipulation. *Adaptive Behavior*, 2003.

[83] Giorgio Metta, L. Natale, S. Rao, and G. Sandini. Development of the mirror system: a computational model. In *Conference on Brain Development and Cognition in Human Infants. Emergence of Social Communication: Hands, Eyes, Ears, Mouths*, Acquafredda di Maratea, Italy, 2002.

[84] Brian Scassellati. Parallel social cognition? In *American Association of Artificial Intelligence Fall Symposium on Parallel Cognitio*, Cape Cod, Massachusetts, 2000.

[85] Brian Scassellati. Theory of mind for a humanoid robot. In *First IEEE/RSJ International Conference on Humanoid Robotics*, September 2000. Best Paper Award.

[86] Brian Scassellati. Theory of mind...for a robot. In *American Association of Artificial Intelligence Fall Symposium on Social Cognition and Action*, Cape Cod, Massachusetts, 2000.

[87] Brian Scassellati. *Biorobotics*, chapter Investigating Models of Social Development Using a Humanoid Robot. MIT Press, 2001.

[88] Brian Scassellati. Discriminating animate from inanimate visual stimuli. *International Joint Conference on Artificial Intelligence*, August 2001.

[89] Brian Scassellati. *Foundations for a Theory of Mind for a Humanoid Robot*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2001.

[90] Brian Scassellati. Foundations for a theory of mind for a humanoid robot. PhD Thesis Defense,Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, May 2001.

[91] Brian Scassellati, Bryan Adams, Aaron Edsinger, and Matthew Marjanovic. Natural tasking of robots based on human cues. Presentation for DARPA Mobile Autonomous Robot Software '01 PI Meeting, March 2001.

[92] Eduardo Torres-Jara. A hand prototype. Internal MIT presentation, 2004.

[93] Eduardo Torres-Jara and Jessica Banks. A simple and scalable force actuator. In *International Symposium of Robotics*, 2004.

[94] Paulina Varchavskaia. Notes on natural language for robots. Presentation for Humanoid Robotics Group, MIT AI Lab, February 2001.

[95] Paulina Varchavskaia. Early pragmatic language development for an infant robot. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2002.

[96] Paulina Varchavskaia, Paul Fitzpatrick, and Cynthia Breazeal. Characterizing and processing robot-directed speech. *IEEE-RAS International Conference on Humanoid Robots 2001*, Nov. 2001.

[97] Paulina Varchavskaya. Behavior-based early language development on a humanoid robot. In *Second International Workshop on Epigenetic Robotics*, Edinburgh, UK, 2002.

[98] Matthew Williamson. *Robot Arm Control Exploiting Natural Dynamics*. PhD thesis, MIT, 1999.