



## Workshop on Cognitive Robots and Systems

### Organizers

Raja Chatila (LAAS-CNRS, France), Henrik Christensen (KTH, Sweden), Ben Kröse (University of Amsterdam, The Netherlands), Roland Siegwart (EPFL, Switzerland).

### Overview

Robots have been applied in a large number of settings over the last decades. However the robots have typically been endowed with rather limited cognitive capabilities in terms of handling general environments and interaction with other agents. Recently there has been significant progress on design of systems that have extended autonomy, methods for acquisition of knowledge, reasoning about environments, tasks and actions and engaging in dialogues with users for task completion. The workshop will review current state of the art on cognitive systems and robotics and will provide an overview of current research on human-robot interaction, architectures, task and skill acquisition, perception and integration.

### Speakers

Rachid Alami (LAAS-CNRS, France)  
Ronald Arkin (Georgia Tech, USA)  
Cynthia Breazeal (MIT, USA)  
Raja Chatila (LAAS-CNRS, France)  
Patrick Jensfelt (KTH, Sweden)  
Rüdiger Dillmann (University of Karlsruhe, Germany)  
Ben Kröse (University of Amsterdam, The Netherlands),  
Benjamin Kuipers The University of Texas at Austin, USA  
Yasuo Kuniyoshi (University of Tokyo, Japan)  
Nicholas Roy (MIT, USA)  
Alan C. Schultz (Naval Research Laboratory, USA)  
Adriana Tapus (University of Southern California, USA)

## **ABSTRACTS**

## **Toward Human-Aware Robot Task Planning**

Rachid Alami  
LAAS-CNRS, Toulouse, France

Human-robot interaction requires explicit reasoning on human environment and on robot capacities to achieve its tasks in a collaborative way with a human partner.

We have devised a decisional framework for human-robot interactive task achievement that is aimed to allow the robot not only to accomplish its tasks but also to produce behaviors that support its engagement vis-a-vis its human partner and to interpret human behaviors and intentions.

Together and in coherence with this framework, we intend to develop and experiment various task planners and interaction schemes that will allow the robot to select and perform its tasks while taking into account explicitly the human abilities as well as the constraints imposed by the presence of humans, their needs and preferences. We present the first results obtained by our "human-aware" task and motion planners and discuss how they can be extended.

## **Behavioral Development for a Humanoid Robot: Towards Life-long human-robot partnerships**

Ronald C. Arkin  
Georgia Tech

This past summer, a significant research effort was conducted at Sony's Intelligence Dynamics Laboratory (SIDL), involving personnel from Georgia Tech, MIT, CMU, Osaka University, and SIDL, working towards the implementation of a theory of designed development for a humanoid robot. This research involves numerous insights gleaned from cognitive psychology (drawn from both new and old theories of behavior) and integrating these techniques into QRIO's architecture with the long-term goal of providing highly satisfying long-term interaction and attachment formation with a human partner. The underlying models used and several results obtained on QRIO are presented.

## **Socio-cognitive skills to support human-robot collaboration**

Cynthia Breazeal  
MIT Media Lab

An important element of cooperative behavior is the ability to understand a partner's actions as well as to infer the goals and other mental states responsible for generating those actions. Simulation Theory argues in favor of an embodied cognition approach

whereby humans reuse parts of their own cognitive structure for not only generating one's own behavior, but also for simulating the mental states responsible for generating that behavior in others. This mental perspective taking skill allows us to recognize, predict, or infer the behavior and intention of others.

This talk presents our simulation-theoretic approach to endowing a robot with mental perspective taking skills and demonstrates its performance in a collaborative teaching/learning scenario. In particular, the ability to interpret demonstrations from the perspective of the teacher plays a critical role in human learning. Robotic systems that aim to learn effectively from human teachers must similarly be able to engage in perspective taking. We present an integrated architecture wherein the robot's cognitive functionality is organized around the ability to understand the environment from the perspective of a social partner as well as its own. The performance of this architecture on a set of learning tasks is evaluated against human data derived from a novel study examining the importance of perspective taking in human learning. Perspective taking, both in humans and in our architecture, focuses the agent's attention on the subset of the problem space that is important to the teacher. This constrained attention allows the agent to overcome ambiguity and incompleteness that can often be present in human demonstrations and thus learn what the teacher intends to teach.

### **An Architecture for Cognitive Robots**

Raja Chatila

LAAS-CNRS, Toulouse, France

A robot system architecture organizes the components of the system at several levels of granularity. These components provide for the functions that the robot will be able to achieve. The architecture describes how these components are linked and interleaved together, how they interact, what information they exchange, and at which temporal rate. The architectural design hence determines the robot's capacities to achieve tasks and to react to events. We propose an architecture that combines four main cognitive capabilities running concurrently, while providing for reactive behavior:

- Spatio-temporal cognition: object, scene and situation learning, recognition and interpretation.
- Decision-making, anticipation and reflexivity: ability to focus attention, to make choices, to plan in advance a course of action, to anticipate the results of these actions or of observed events, to reflect on own abilities.
- Learning: improving abilities with experience, acquiring new abilities, building new knowledge from acquired information.
- Interaction and communication: understanding multi-modal commands (taskability of the robot) and expressing intentions.

This architecture is being implemented in different instances in the framework of the Cogniron project.

**Cognitive Systems for Cognitive Assistants**

Henrik I Christensen & Patric Jensfelt  
Center for Autonomous Systems  
Royal Institute of Technology  
Stockholm, Sweden

Despite impressive progress in many specific sub-topics in AI and Cognitive Science, work on building integrated cognitive systems moves slowly. Most systems able to perform complex tasks that humans and other animals can perform easily, for instance robot manipulators, or intelligent advisers, have to be very carefully crafted, normally their field of expertise is very narrow, and they are difficult to extend. Whatever intelligence they have could be described as "insect-like", with very little flexibility or self-understanding.

Part of the reason for this is that over the last few decades research has become highly fragmented: with many individuals and research teams focusing their efforts on narrowly defined problems, for instance in vision, or learning, or language processing, or problem solving, or mobile robotics. The CoSy project aims to overcome these limitations by using ideas from several relevant disciplines to investigate an ambitious vision of a highly competent robot/human companion, combining different capabilities in a coherent manner. The work is centered around two main types of objectives concerned with theory and implementation.

CoSy involves studies of theoretical models, at different levels of abstraction, regarding requirements, architectures, forms of representation, kinds of ontologies, types of reasoning, kinds of knowledge, and varieties of mechanisms relevant to embodied, integrated, multi-functional intelligent systems. The results should be useful both for enhancing scientific understanding of naturally occurring intelligent systems (e.g. humans) and for the design of artificial intelligent systems.

In addition CoSy implements a succession of increasingly sophisticated working systems demonstrating applications of parts of the theory, e.g. in a robot capable of performing a diverse collection of tasks in a variety of challenging scenarios, including various combinations of visual and other forms of perception, learning, reasoning, communication and goal formation. Initially, two main kinds of robots are investigated both of which will learn from and interact with human teachers. One system does spatial modelling while the other does interpretation / replication of manipulative actions.

This presentation provides an overview of the CoSy project, the main objectives, theoretical results and implementations results from the first year of operation. The presentation will cover language/dialogue systems, fault-detection, planning, representations, object categorisation and integration.

## **Tracking and interpretation of human body motion**

Rüdiger Dillmann, Steffen Knoop  
Institute of Computer Science and Engineering (CSE)  
University of Karlsruhe  
Germany  
Email: {knoop,dillmann}@ira.uka.de

Robots that are meant to cooperate closely with humans, and especially with untrained persons which are not familiar with the domain of robotics, need a deep understanding of the intentions, activities, actions and movements of their interaction partner.

Many activities can be recognized by observing movements and posture of the human body. Additional information for determining a person's current state and activities can be derived from the environment. This information is often called context. Together, body motion and context provide in many situations enough information to derive the person's current activity or activities.

For estimation of movements as well as gestures and activities of a human interaction partner, we divide the process into two tasks: (1) Realtime tracking of the human body pose over time, giving a trajectory of the whole body motion, and (2) interpretation of the observed movements to detect and recognize gestures and basic activities.

We present a tracking system called VooDoo for 3d tracking of human body movements based on a 3d human body model and the Iterative Closest Point (ICP) algorithm. The proposed approach is able to incorporate raw data from different input sensors, as well as results from feature trackers in 2d or 3d. Typical sensor sources are a 3D time-of-flight sensor point clouds, color based feature segmentation (e.g. skin color), and laser range finder measurements. All input data is processed within the same model fitting step by projecting the measurements to 3d model space. For cases where no 3d data is available, an alternative purely vision-based particle filtering approach will be presented. Both systems have been implemented and run in realtime at appr. 15-20 Hz.

Our approach for human activity recognition in Human-Robot Interaction relies on the observed trajectory of the articulated human body. This intrinsic information is used together with extrinsic context features to recognize and classify human activities.

## **From sensor data to human spatial concepts: an appearance-based approach**

Ben Kröse, Zoran Zivkovic and Olaf Booij,  
University of Amsterdam

For a natural communication with humans, a cognitive robot must be able to relate its sensor reading to human spatial concepts such as: "the living room", "behind TV" etcetera. Spaces (or objects) can be represented by a geometric (3D) model of the space/object, or represented by an appearance model. In our lab we work on appearance based algorithms where the sensor reading can be related to the human

concepts more directly. The appearance based approaches can be used also for navigation. In the early work the appearance based representation consisted of a set of sensor readings from known locations and PCA was used to reduce the size of the data set. In the recent work the locations of the sensor readings are not needed. The appearance based representation is constructed by extracting the similarities and relations between the sensor readings, for example we relate omnidirectional images using local features epipolar geometry constraints. Clusters in this representation correspond to convex spaces in an indoor environment separated by walls and other barriers. By interacting with humans such clusters can be labeled with human concepts and refined in order to arrive to a space representation appropriate for communicating with humans.

### **The Hybrid Spatial Semantic Hierarchy: Factoring the Mapping Problem**

Benjamin Kuipers  
Computer Science Department  
University of Texas at Austin

A map is a description of an environment allowing an agent, human or robot, to plan effective actions within the environment, particularly motion from place to place. In a complex, large-scale environment, the structure as a whole cannot be observed by the robot's sensors from a single vantage point, so the robot must create a map from observations gathered over time and travel.

Purely metrical approaches to environmental mapping typically define a single global frame of reference for the map to be created, and then do probabilistic inference from observations to determine cell occupancy or feature poses. Global consistency problems rapidly arise, particularly when closing large loops of travel in the environment. The fundamental problem is representational: loop-closing hypotheses should be represented as alternative topological structures for the map, not as alternatives in the vastly larger space of metrical maps.

Our Hybrid Spatial Semantic Hierarchy approach combines topological and metrical representations, and applies them to both large-scale and small-scale space. The Hybrid SSH factors the mapping problem into three natural sub-problems. The first, building metrical maps of local, small-scale environments, is solved effectively by existing SLAM methods. The second, building a global topological map of large-scale space, is solved as a search over the space of topological maps consistent with exploration experience. The third, building the global metrical map, is solved efficiently and accurately once we have the topological skeleton to build on. The Hybrid SSH specifies the abstractions that link these different representations, the synergies they support, and the (weak) assumptions they make about the robot's sensory and motor capabilities.

**Body Image Acquisition for Robotic Tool Use and Imitation**

Yasuo Kuniyoshi

Interfaculty Initiative in Information Studies &  
Dept. of Mechano-Informatics, School of Information Sci. & Tech.  
The Univ. of Tokyo, Japan.

**Model-Uncertainty Planning and Control**

Nicholas Roy,

Department of Aeronautics and Astronautics,  
MIT

Decision making in uncertain and incomplete models is an essential capability of robots operating in natural, dynamic domains. Separating statistical model learning and planning into two distinct processes simplifies both problems, but building complete and accurate models becomes increasingly difficult as the world increases in size and complexity, and the planner becomes brittle to model errors. For example, a mobile manipulator grasping an object on a table should not depend solely on the prior existence of an exact and detailed model of the table and object. I will describe some approaches to model-uncertainty control using reinforcement learning in the information space of model distributions to generate plans that minimize an expected cost with respect to the distribution over models.

## **Using Computational Cognitive Models to Build Better Human-Robot Interaction**

Alan C. Schultz and J. Gregory Trafton  
Navy Center for Applied Research In Artificial Intelligence  
Naval Research Laboratory, Washington DC 20375  
{schultz,trafton}@aic.nrl.navy.mil

### **Introduction**

We propose an approach for creating more cognitively capable robots that can interact more naturally with humans. Through analysis of human behavior, we build computational cognitive models of particular high-level human skills that we have determined to be critical for good peer-to-peer collaboration and interaction. We then take these cognitive models, place them on the robot, and allow the robot to use the models to make decisions.

### **Cognitively enhanced intelligent systems**

We hypothesize that adding computational cognitive reasoning components to intelligent systems such as robots will result in three benefits:

Most if not all intelligent systems must interact with humans, who are the ultimate users of these systems. Giving the system cognitive models can enhance the human-system interface by allowing more common ground in the form of cognitively plausible representations and qualitative reasoning. For example, mobile robots generally use representations such as rotational and translational matrixes to represent motion and spatial references. However, this is not a natural mechanism for humans, and results in additional computations to translate between these and the qualitative spatial reasoning used by humans. By using cognitive models, reasoning mechanisms and representations, we believe that we can yield a more effective and efficient interface.

Since the resulting system is interacting with the human, giving it behaviors that are more natural and compatible with the human can also result in more natural interactions between the human and the intelligent system. For example, mobile robots that must work collaboratively with humans can actually result in less effective interactions if its behaviors are alien or non-intuitive to the human. By incorporating cognitive models, we can develop systems whose behavior is more expected and natural.

One key interest is in measuring the performance of intelligent systems. We propose that an intelligent system that is cognitively enhanced can be more directly compared to human level performance. Further, if cognitive models of human performance have been developed in creating the intelligent system, we can directly compare the intelligent systems behavior and performance in the task to the human subject behavior and performance.

## Hide and Seek

Our foray into this area started when we were developing computation cognitive models of how young children learn the game of hide and seek (Trafton et al. 2005, Trafton et al. 2006). The purpose was to enable our robots to use human-level cognitive skills to make the decisions about where to look for people or things hidden by people. The research resulted in a hybrid architecture with a reactive/probabilistic system for robot mobility (Schultz, Adams & Yamauchi, 1999), and a high-level cognitive system based on ACT-R (Anderson & Lebiere, 1998) that made the high-level decisions for where to hide or seek (depending on which role the robot was playing).

While this work was interesting in its own right, the system led us to the realization that the ability to do perspective taking was a critical cognitive ability for humans, particularly when they want to collaborate.

## Spatial perspective taking

To determine just how important perspective and frames of reference were in collaborative tasks in shared space (and also because we were working on a DARPA funded project to move these capabilities to the NASA Robonaut), we analyzed a series of tapes of two astronauts and a ground controller training in the NASA Neutral Buoyancy Tank facility for an assembly task for Space Station mission 9A. We performed a protocol analysis of several hours of these tapes focusing on the use of spatial language and commands from one person to another. We found that the astronauts changed their frame of reference (as seen during their dialog) approximately every other utterance. As an example of how prevalent these changes in frame of reference are, consider this following utterance from ground control:

“... if you come *straight down* from where you are, uh, and uh, kind of peek *down under the rail* on the *nadir side*, by *your right hand*, almost *straight nadir*, you should see the...”

Here we see five changes in frame of reference (highlighted in italics) in a single sentence! These rates in the change of reference are consistent with work by Franklin, Tversky, & Coon, 1992. In addition, we found that the astronauts had to take other perspectives, or forced others to take their perspective, about 25% of the time (Trafton et al., under review). Obviously, the ability to handle changing frames of reference and being able to understand spatial perspective will be a critical skill for robots such as NASA Robonaut and, we would argue, any other robotic system that needs to communicate with people in spatial contexts (i.e., any construction task, direction giving, etc.).

## Models of perspective taking

Imagine the following task, as illustrated in Figure X. An astronaut and his robotic assistant are working together to assemble a structure in shared space. The human, who can see one wrench, says to the robot, “Pass me the wrench.” Meanwhile, from the robot’s point of view, two wenches are visible, while the human has a partially occluded view and can only see one wench. What should the robot do? Evidence suggests that

humans, in similar situations, will pass the wench that they know the other human can see (Clark, 1996) since this is a jointly salient feature.

We developed two models of perspective taking that could handle the above scenario in a general sense. The first approach used the ACT-R/S system (Harrison & Schunn, 2002) to model perspective taking using a cognitively plausible spatial representation. The second approach used Polyscheme (Cassimatis et al., 2004) and modeled the cognitive process of mental simulation; humans tend to mentally simulate situations in order to resolve problems.

Using these models we have demonstrated a robot being able to solve problems similar to the wench problem.

### **Future work**

We are now exploring other human cognitive skills that seem important for peer-to-peer collaborative tasks and that are appropriate for building computational cognitive models for adding to our robots. One new skill we are considering is non-visual, high-level focus of attention. This skill helps to focus a person's attention to appropriate parts of the environment or situations based on the current environment, task, expectations, models of other agents in the environment and other factors. Another human cognitive skill we are considering involves the role of anticipation in human interaction and decision making.

### **Conclusion**

It is clear that if humans are to work as peers with robots in shared space, the robot must be able to understand the natural human tendency to use different frames of reference and to take the human's perspective. To create robots with these capabilities, we propose using CCMs as opposed to more traditional programming paradigms for robots. First, a natural and intuitive interaction results in reduced cognitive load. Second, more predictable behavior engenders trust. Finally, more understandable decisions allow the human to recognize and more quickly repair mistakes.

We believe that using computational cognitive models will give our robots the cognitive skills necessary to interact more naturally with humans, particular in peer-to-peer relationships.

### **References**

J. G. Trafton, Alan C. Schultz, N. L. Cassimatis, L. Hiatt, D. Perzanowski, D. P. Brock, M. Bugajska, and W Adams, (2005). "Using Similar Representations to Improve Human-Robot Interaction," Agents and Architectures (Ed. Ron Sun), Erlbaum, 2005.

J. G. Trafton, Alan C. Schultz (2006). "Children and robots learning to play hide and seek," ACM conference on Human-Robot Interaction, March 2006.

N. Cassimatis, J. G. Trafton, M. Bugajska, and A. C. Schultz (2004). "Integrating Cognition, Perception, and Action through Mental Simulation in Robots." *Robotics and Autonomous Systems*, 49(1-2), Elsevier, Nov. 2004, pp. 13-23.

Trafton, J. G., Cassimatis, N. L., Brock, D. P., Bugajska, M. D., Mintz, F. E., & Schultz, A. C. (2005). "Enabling effective human-robot interaction using perspective-taking in robots," *IEEE Transactions on Systems, Man, and Cybernetics---Part {A}: Systems and Humans*, 35(4), 460-470.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, 20(5), 507-518.

Schultz, A., Adams, W., and Yamauchi, B. (1999). "Integrating Exploration, Localization, Navigation and Planning Through a Common Representation," *Autonomous Robots*, 6(3), June, 1999.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Erlbaum.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Harrison, A. M., & Schunn, C. D. (2002). ACT-R/S: A computational and neurologically inspired model of spatial reasoning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the Twenty Fourth Annual Meeting of the Cognitive Science Society* (pp. 1008). Fairfax, VA: Lawrence Erlbaum Associates.

## **From Human-like Robot Navigation to Human-Robot Interaction**

Adriana Tapus

Robotics Research Lab / Interaction Lab,  
University of Southern California, Los Angeles, USA  
[tapus@robotics.usc.edu](mailto:tapus@robotics.usc.edu)

The undeniable trend of research in robotics is to endow robots with the capability of understanding the world we are in, thus permitting them to help us and to be a part of our lives. An ideal companion-robot should be designed to feature sufficiently complex cognitive capabilities permitting it to understand and to interact with the environment, to exhibit social behaviour, and to focus its attention and communicate with people. In all our daily activities, the natural surroundings that we inhabit play a crucial role. Therefore, first the problem of perception, spatial cognition and topological navigation for a mobile robot will be addressed. All these elements are needed in order to obtain a robust and reliable framework for navigation. A new paradigm "fingerprints of places" to solve these problems is introduced in this work. Fingerprints of places (i.e. circular list of significant features around the robot) have been proven to be a very promising approach towards effective place characterization and hence environment modeling. The fingerprint approach, by combining the information from all sensors available to the robot, reduces perceptual aliasing and improves the distinctiveness of places. The construction of a

topological mapping system is combined with localization, both relying on fingerprints of places, in order to perform Simultaneous Localization and Mapping (SLAM). Our fingerprint based approach yields a consistent and distinctive representation of the environment and is extensible in that it permits spatial cognition beyond just pure navigation. This enables navigation of an autonomous mobile robot in a structured environment without relying on maps given a priori, without using artificial landmarks and by employing a semantic spatial representation that allows a more natural interface between humans and robots.

With a view of having robots as companion of humans, we are motivated towards developing a knowledge representation system along the lines of what we know about us. Therefore, in order to build a natural solution for human-robot interaction many questions and issues should be addressed, including: 'What are the circumstances in which people accept a robot in their environment?', 'How can the interaction design ensure safety, by allowing users to easily stop the robot or change its actions?', 'How can friendly and familiar interaction models with a robot be designed?', 'What modes of communication should be employed?', 'How to model the behavior as a function of the personality of the user?', 'How to integrate the *a priori* knowledge about the users in the robotic system?', 'How can the robot's perception, competence, and awareness of the world be represented in a form accessible and understandable to a non-technical user?', 'What models from psychology, cognitive science, and social science can be utilized?', 'What's the role of embodiment ?'.

This work details the complete methodology for human-like navigation, the experiments conducted, the results obtained, the problems that have been solved and the ongoing research directed towards human-robot interaction.