Open access • Journal Article • DOI:10.1109/TAMD.2012.2199754

Towards a Platform-Independent Cooperative Human Robot Interaction System: III An Architecture for Learning and Executing Actions and Shared Plans — Source link

Stephane Lallee, Ugo Pattacini, Séverin Lemaignan, Alexander Lenz ...+13 more authors

Institutions: French Institute of Health and Medical Research, Istituto Italiano di Tecnologia, Max Planck Society, Harvard University

Published on: 01 Sep 2012 - IEEE Transactions on Autonomous Mental Development (IEEE)

Topics: Humanoid robot, Robot kinematics, Software portability, Computer-supported cooperative work and Human–robot interaction

Towards a Platform-Independent Cooperative Human-Robot Interaction System: II. Perception, Execution and Imitation of Goal Directed Actions

Stephane Lallée, Ugo Pattacini, Jean David Boucher, Séverin Lemaignan, Alexander Lenz, Chris Melhuish, Lorenzo Natale, Sergey Skachek, Katharina Hamann, Jasmin Steinwender, Emrah Akin Sisbot, Giorgio Metta, Rachid Alami, Matthieu Warnier, Julien Guitton, Felix Warneken, Peter Ford Dominey

Abstract— If robots are to cooperate with humans in an increasingly human-like manner, then significant progress must be made in their abilities to observe and learn to perform novel goal directed actions in a flexible and adaptive manner. The current research addresses this challenge. In CHRIS.I [1], we developed a platform-independent perceptual system that learns from observation to recognize human actions in a way which abstracted from the specifics of the robotic platform, learning actions including "put X on Y" and "take X". In the current research, we extend this system from action perception to execution, consistent with current developmental research in human understanding of goal directed action and teleological reasoning. We demonstrate the platform independence with experiments on three different robots. In Experiments 1 and 2 we complete our previous study of perception of actions "put" and "take" demonstrating how the system learns to execute these same actions, along with new related actions "cover" and "uncover" based on the composition of action primitives "grasp X" and "release X at Y". Significantly, these compositional action execution specifications learned on one iCub robot are then executed on another, based on the abstraction layer of motor primitives. Experiment 3 further validates the platform-independence of the system, as a new action that is learned on the iCub in Lyon is then executed on the Jido robot in Toulouse. In Experiment 4 we extended the definition of action perception to include the notion of agency, again inspired by developmental studies of agency attribution, exploiting the Kinect motion capture system for tracking human motion. Finally in Experiment 5 we demonstrate how the combined representation of action in terms of perception and execution provides the basis for imitation. This provides the basis for an open ended cooperation capability where new actions can be learned and integrated into shared plans for cooperation. Part of the novelty of this research is the robots' use of spoken language understanding and visual perception to generate action representations in a platform independent manner based on physical state changes. This provides a flexible capability for goal-directed action imitation.

I. Introduction

For embodied agents that perceive and act in the world, there is a strong coupling or symmetry between perception and execution which is constructed around the notion of goal directed action. Hommel et al [2] propose a philosophy for the cognitive mechanisms underlying perception and action – the Theory of Event Coding. According to this theory, the stimulus representations underlying action perception, and the sensorimotor representations underlying action are not coded separately, but instead are encoded in a common representational format. In this context it has now become clearly established that neurons in the parietal and the premotor cortices encode simple actions both for the execution of these actions as well as for the perception of these same actions when they performed by a second agent [3]. This research corroborates the emphasis from behavioral studies on the importance of the goal (rather than the details of the means) in action perception [4].

Within a sensorimotor architecture a number of benefits derive from such a format, including the direct relation between action perception and execution that can provide the basis for imitation. This is consistent with our previous research in the domain of robot perception and action in the context of cooperation ([5, 6]). The current research extends our previous work on the learning of composite actions by exploiting this proposed relation between action execution and perception. Part of the novelty of the current research is that the action repertoire is open: the robot can learn new actions in both dimensions of perception and execution. The learned actions take arguments including agent, object and recipient. Maintaining this symmetry of action perception and execution lays the framework for imitation and the use of imitation in cooperation [5, 6].

We look to human development to extract requirements on how to implement such an action representation. In this context, two important skills for infants are the ability to detect an action as being goal directed and to determine its

Manuscript received March 15, 2011. This work was fully supported by European FP7 ICT project CHRIS). Stephane Lallee, Jean-David boucher and Peter Ford Dominey are with the Stem Cell & Brain Research Institute, INSERM U846, Bron, France. (stephane.lallee@inserm.fr). Severin Lemaignan, Emrah Akin Sisbot and Rachid Alami are with LAAS, CNRS, Toulouse, France. Alexander Lenz, Chris Melhuish and Sergey Skachek are with BRL, Bristol, United Kingdom. Ugo Pattacini,Lorenzo Natale and Giorgio Metta are with IIT, Genoa, Italy. Jasmin Steinwender and Katharina Hamann are with Max Planck Institute, Leipzig, Germany. Felix Warneken is with Harvard University, Cambridge, USA

agency. Studies of infant action perception [4, 7] have led to the extraction of a core set of conditions which allows the infant to identify goal-directed actions. In the current research, we implement in our system the ability to address aspects of these human requirements both in terms of perception (detect and represent *salient actions effects*) and execution (ability to achieve a goal through an action using *equifinal variations*). We demonstrate how those capabilities can be used by the robot to imitate or mirror human actions (which involve both recognition and execution) in a way that should match the human requirements for goal attribution.

Learning by imitation is a major area of research in robot cognition today [8-12]. Our novel contribution to this domain is the encoding of action in terms of perceptual state changes and composed motor primitives that can achieve these state changes, in a manner that allows the robot to learn new actions as perception – execution pairs, and then use this knowledge to perceive and imitate. These actions can take several arguments, e.g. `AGENT put the OBJECT on the RECIPIENT`. This allows for the generalization of learned actions to entirely new contexts, with new objects and agents. In our long-term research program, this provides the basis for learning to perform joint cooperative tasks purely through observation.

II. CONTEXT: GOAL DIRECTED ACTIONS

A. *Goals attribution requirements*

Studies of human infants [4, 13-15] indicate that their ability to determine the goal of an action begins to develop between 6 and 9 months, demonstrated by the ability of infants to encode behaviors such as a hand grasping for an object as being directed at the goal-object rather than encoding the hand's specific movement. An important issue that has been discussed within the field is the difference between actions that are familiar to the infant and more unfamiliar actions which may not include human features (like a robotic gripper grasping a toy). Woodward [14] initially argued that only observed actions that the infant is able to execute herself are represented as goal-directed. However later studies [4, 7] demonstrated that indeed infants are able to attribute goal directedness for novel actions early assuming two conditions: first the action has to produce a salient effect on the world state (like the motion from one place to another). The second condition is that the agent is able to achieve the same state change in different ways (such as avoiding an obstacle instead of using a straight trajectory), in other words the action is demonstrated to possess equifinal variations.

B. *Implementing those requirements*

Our implementation of action, both in the context of perception from CHRIS.I [1] and execution is based on actions as state changes. One of the strong implications of this is the equifinality of action. That is, the same action "put the box on the toy" may be realized in a variety of ways (with one hand, or the other) but with the equivalent final outcome, one of the key characteristics that allow action to be considered goal directed. If the robot is able to demonstrate equifinal means of achieving his actions, then humans may be more likely to attribute a goal to them. This assumption has been shown to be true in infants [4, 16] and would need to be tested on adults, however assuming the fact that all our teleological system seems to be built on those core capabilities it is likely that a benefactor effect could be found also on adults.

In our action recognition system [1] we exploited Mandler's [17] suggestion that the infant begins to construct meaning from the scene based on the extraction of perceptual primitives. From simple representations such as contact, support and attachment [18] the infant could construct progressively more elaborate representations of visuospatial meaning. In this context, the physical event "collision" can be derived from the perceptual primitive "contact". Kotovsky & Baillargeon [19] observed that at 6 months, infants demonstrate sensitivity to the parameters of objects involved in a collision, and the resulting effect on the collision, suggesting indeed that infants can represent contact as an event predicate with agent and patient arguments.

In this paper we describe an evolution of the action recognition system described in [1]. This new system is still based on sequences of perceptual event primitives (visibility, motion, contact), however those primitives are now represented in terms of the impact they have on the world state. Primitives can be queued and their effects added so that a sequence of them will be a way to reach an end state from an initial state. If a sequence produces no change in the world state, then it will not be taken into account by the system, which mimics the ability of children to emphasis actions that produce a salient effect on the world. This rejection of "useless" actions allow the system to be more stable: for example an object which appears and then disappears quickly may be only a false recognition of the perceptual system.

These requirements are implemented on both the perceptual and executive components of the system. In CHRIS.I [1] we presented a system architecture for cooperation. Here we zoom in on the action related components which handle the complete link from perception to motor commands in term of actions.

III. Experimental Platforms

A crucial aspect of our research is that the architecture should allow knowledge acquired on one robot to be used on physically distinct platforms. In the current study this is demonstrated using two different version of the iCub platform in Lyon France, and Genoa Italy, respectively, and the Jido robot in Toulouse, France.

The iCub [20] is an open-source robotic platform shaped as three and a half year-old child (about 104cm tall), with 53 degrees of freedom distributed on the head, arms, hands and legs. The head has 6 degrees of freedom (roll, pan and tilt in the neck, tilt and independent pan in the eyes). Three degrees of freedom are allocated to the waist, and 6 to each leg (three, one and two respectively for the hip, knee and ankle). The arms have 7 degrees of freedom, three in the shoulder, one in the elbow and three in the wrist. The iCub has been specifically designed to study manipulation, for this reason the number of degrees of freedom of the hands has been maximized with respect to the constraint of the small size. The hands of the iCub have five fingers and 19 joints. All the code and documentation is provided open source by the RobotCub Consortium, together with the hardware documentation and CAD drawings. The robot hardware is based on high-performance electric motors controlled by a DSP-based custom electronics. From the sensory point of view the robot is equipped with cameras, microphones, gyroscopes, position sensors in all joints, force/torque sensors in each limb.

While both iCubs are instances of the iCub, they are distinct in the implementation of motor control as the iCubGenoa01 is equipped with force sensors that allow force control; the iCubLyon01 is only controlled in velocity and position modes. Thus, the essential role of the motor primitive pool as the common abstraction layer across robots is maintained. Jido, on the other hand is an entirely different robot, which allows us to truly explore the platform independence of our system.

Jido is a fully-equipped mobile manipulator that has been constructed in the framework of Cogniron (IST FET project: www.cogniron.org). Jido, a MP-L655 platform from Neobotix, is a mobile robot designed to interact with human beings. It is presented on figure 3. Jido is equipped with: (i) a 6-DOF arm, (ii) a pantilt unit system at the top of a mast (dedicated to human-robot interaction mechanisms), (iii) a 3D swissranger camera and (iv) a stereo camera, both embedded on the pan tilt unit, (v) a second video system fixed on the arm wrist for object grasping, (vi) two laser scanners, (vii) one panel PC with tactile screen for interaction purpose, and (viii) one screen to provide feedback to the robot user. Jido has been endowed with functions enabling to act as robot companion and especially to exchange objects with human beings. So, it embeds robust and efficient basic navigation and object recognition abilities.

IV. The CHRIS architecture – Focus on action

In order to be platform-independent, action representation is abstracted from platform-specificities at the lowest level possible. An overview of the CHRIS architecture in this context is presented in Figure1.

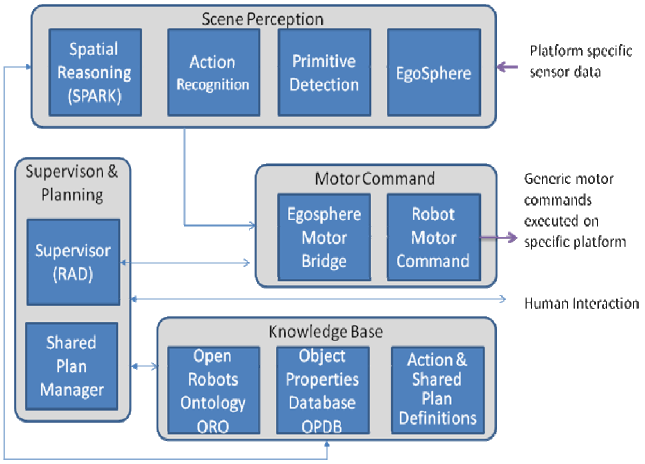

Figure 1: CHRIS Architecture. Arrows represent the flow of information (data, commands), which are transported over the network via YARP. Perceptual information enters Scene Perception. Object positions from Egosphere are processed by Primitive Recognizer and Action Recognizer for learning and recognition, and enter SPARK for inference of spatial relations which are stored in ORO. Shared Plan Manager links perceptual and exectutive action representations and plans. Supervisor manages HRI, the learning of new action execution, and verificiation from ORO that execution preconditions hold.

A. *Abstraction of Action Perception and Execution*

Two layers of abstractions are required in order to have a platform independent architecture: perceptual and motor. Both of them rely on the Egosphere module.

1) *Scene Perception*

The first layer of abstraction between the sensory perception systems and the higher level cognitive architecture and motor control elements is formed at the level of the Egosphere which serves as a fast, dynamic, asynchronous storage of object positions and orientations. The object positions are stored in spherical coordinates (radius, azimuth and elevation) and the object orientation is stored as rotations of the object reference frame about the three axes (x,y,z) of a right-handed Cartesian world-frame system. The origin of the world frame can be chosen arbitrarily and, for our experimental work, we located it at the centre of the robot's base-frame. Other stored object properties are a visibility flag and the objectID. The objectID is a unique identifier of an object which acts as a shared key across several databases (see [1] for details). The robot-specific 3D perception system adds objects to the Egosphere when they are first perceived, and maintains position, orientation or visibility of these objects over time. Modules requiring spatial information about objects in the scene can query the Egosphere. The Egosphere is implemented in C++ as a client-server system using the YARP infrastructure. Software modules requiring access to the Egosphere include a client class which provides methods like addObject(), setObject(), getObject() or getNumberOfObjects(), etc. The Egosphere is thus a convenient abstraction layer. With increasing complexity of human-robot interaction tasks during the course of our research, we will add further complexity (human focus of attention, confidence, timeliness etc.) whilst preserving modularity. This is exemplified by the spatial reasoning (e.g. visibility by line of sight) provided by Spark. Within the Jido platform-independent component, the functionality of the EgoSphere is preserved within Spark.

2) *Perceptual Primitives, Events and State Changes*

The action recognition capability is based on the extraction of meaningful primitive events from the flow of object positions and visibilities represented in the Egosphere and Spark. Again we based our system findings from developmental psychology. We implemented perceptual primitives similar to those described in [21-25]. We have previously used this primitives based approach in [26, 27] and we identified a core set of primitive events that are simple and provide a solid basis for action construction. There are six primitive event divided in three categories:

- Visibility (object appears or disappears)
- Motion (object starts or stops moving)
- Contact (contact made or broken between 2 objects)

Each of these primitive event is coded in terms of the state change it effects on the world (e.g: if an object appears, *visibility(object)* will be added to the world state). The Primitive Recognizer extracts those 6 primitives by constantly monitoring the Egosphere. It then broadcasts the detected events to the Action Recognizer.

3) *Motor Primitives*

The current research extends this notion of compositionality for action perception from CHRIS.I [1] to action execution. As for the perceptual system, the action execution system requires a suitable abstraction that provides a platform independent interface to the robot motor capabilities. Motor primitives rely on the idea that complex motor tasks may be achieved by the combination of simple parameterized controllers we call primitives. This framework is consistent with studies of biological motion [28], which demonstrate that motion of biological beings is achieved by high level motor commands triggering a sequence of motor primitives leading finally to an effective motion of the muscles. Using hierarchies of primitives for control in robotics is becoming a widely used method [29-36]. In our approach, what we call a Motor Primitive is already a symbolic action. The implementation of those actions is robot specific, what is important is that all robots share the same motor interface, as a pool of Motor Primitives. In the current system the primitives that are implemented on the robot are:

- Grasp (object)
- Release (location)
- Touch (object)
- Look-At (object)

We do not claim the completeness of this pool for all possible interactions, but these primitives were sufficient in the context of robot and human interaction through manipulation of objects on a table. The arguments for these primitives are objects whose Cartesian coordinates are recovered from the Egosphere.

B. *Action Representation*

The concept of Action and its representation is at the center of our architecture. Inspired by the perception-execution symmetry [2] we impose the requirement that the same data structure shall accommodate both the perceptual and executive components of action. It also includes teleological information, that is, the state changes that are induced by that action.

1) *Action Representation for Perception*

Our representation of action started with a purely perceptual definition [1, 6, 37]. Specifically the Action Recognizer module is constantly monitoring the flow of perceptual primitives sent by the Primitive Recognizer module. We make the assumption that two actions will be separated by a temporal delay, so we can use this delay to segment meaningful sequences of primitives. When such an independent sequence is detected, it is tagged as being a potential action which is then evaluated by the recognition process. The action data structure is similar to that for events since actions are composed of primitive events, and both produce a salient change (or changes) in the world state. The Action Recognizer stores a list of all the known actions and compares them with the incoming potential actions. All the primitives contained in the received sequence are added so that the global world state change of this sequence is obtained, then if a known action creates the same change in the environment it is recognized as being the observed action. We have to stress the fact that this “world change” is argument independent: if the system has learnt an action *cover(object A, object B)* then it will recognize a *cover(toy,box)* as well as a *cover(bowl, plate).*

Actions possess characteristics in addition to those of event primitives. The state change produced by an event primitive is called post-condition, because it is applied after the primitive occurred. In addition to post-conditions an action has pre-conditions which can either allow or prevent it to occur (for example covering the bowl needs the bowl to be visible and uncover the bowl needs the bowl to be covered). Those pre/post conditions are a useful mechanism that allows forward/backward chaining and finally teleological reasoning (see [37] for more details about this aspect). Actions also contain a field describing the executing agent. Agency detection is based on motion primitives associated with human hands that are detected using the Kinect device which provides information about human hands to the Egosphere (see below).

2) *Action Representation for Execution*

In order to bridge the gap between perception and execution, the Shared Plan Manager module combines motor representations with perceptual representations of action. While we currently address the learning of single actions as the simplest motor plans, the system is designed to naturally extend to more complex shared plans, based on our earlier work [6].

When the user asks the robot to perform an action the Shared Plan Manager searches for a plan with that name. If no such plan is found, then the Shared Plan Manager asks the user to enumerate the motor primitives (described above) that constitute that action.

The system can thus learn to perform complex actions such as put `the box on the toy` as a composite sequence of `grasp box, release box on toy`. We implement a form of argument binding so that this newly learned action can generalize across all objects. That is the robot can then perform the action `put the toy on the table`.

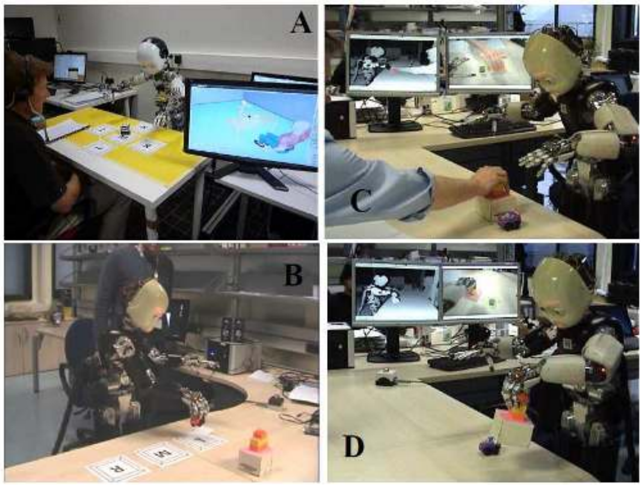

Figure 2: Experiments on iCubLyon01 and iCubGenoa01. A. Experiments 1 and 2 where human teaches robot new actions. Note in right foreground the representation of the spatial environment in SPARK. B. Replication of actions learned in Lyon with iCubLyon01 transferred to iCubGenoa01 in Genoa. C. Human demonstrates the "cover the toy with the box" action, and the iCubGenoa01 recognizes and imitates that action.

C. *Supervision*

Action perception and execution are coordinated by the HRI Supervisor. The Supervisor manages spoken language interaction with the CSLU Toolkit [38] Rapid Application Development (RAD) state-based dialog system which combines state-of-the-art speech synthesis (Festival) and recognition (Sphinx-II recognizer) in a GUI programming environment. Our system is thus state based, with the user indicating the nature of the current task (including whether he wants interact in the context of action recognition, execution or imitation tasks). In each of these subdomains, the user can then indicate that he is ready to show the robot a new example and the robot will attempt to recognize, perform or learn what is shown.

A principal function of the Supervisor is to verify that preconditions for action execution are met before the execution is initiated. This primarily concerns the constraint that objects to be manipulated should be visible. This information is computed by the SPARK (Spatial Reasoning and Knowledge) module and made available to the system in ORO (the Open Robot Ontology) which provides central component of the Knowledge base of the system. See CHRIS.I [1] for details.

V. Experiments

A. *Experiment 1- Completing Perception with Execution*

In CHRIS.I we demonstrated a capability to learn to recognize actions including `take` and `put`. Here we first demonstrate how these action definitions can be completed with the execution component.

H: Put the toy on the left
R: I don't know how to put.
H: Grasp the toy.
R: Grasping the toy.
H: Release left
R: Releasing left
H: Finish learning.

Based on this learning we then demonstrated that the acquired execution knowledge could generalize to new instances of the action. We demonstrated that the robot correctly performed the command to put the box in the middle. This is illustrated in Fig 2A. In order to demonstrate that this knowledge could be exploited on a different robot, the learned definitions were shared via the SVN repository. Figure 2B illustrates the iCubGenoa01 using action definitions acquired in Lyon in order to perform the take and put actions.

B. *Experiment 2- Learning New Actions*

This experiment tests the ability of the system to learn new actions, both in terms of perception and execution. Here we focus on two actions which are `cover X with Y`, and `uncover X with Y`. We chose these actions as they will provide the basis for future work in shared planning for cooperation.

H: Cover the toy with the box.
R: I do not know how to cover.

H: Grasp the box.
R: Grasping the box.
H: Release the box on the toy.
R: Releasing the box on the toy.
H: Finish learning.

This dialog fragment illustrates how the system can acquire new sequences of action primitives in order to learn new composite actions. Here, "cover X with Y" is learned as the concatenation of grasp X and release X at Y. We demonstrated this same concatenative learning for the actions, put, take, cover and uncover. Note that put and cover have similar definitions, with reversed ordering of the arguments, demonstrating the flexibility of the argument binding capability.

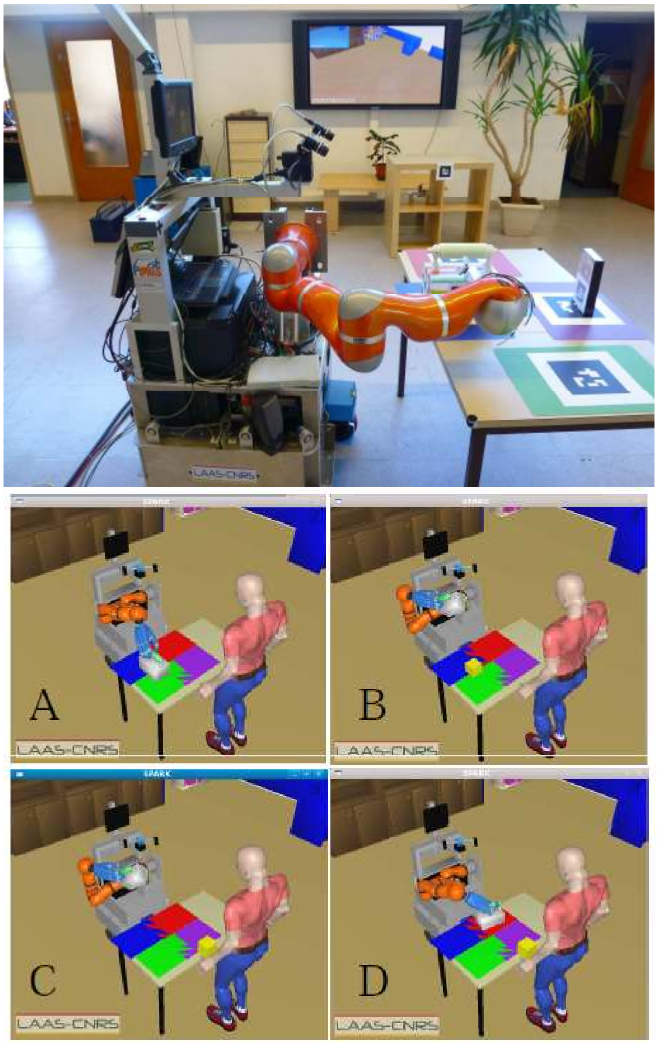

Figure 3: Above - Experimental platform Jido. The action of taking the box and putting it on the red-mat (cover X with Y) that was learned on iCubLyon01 was successfully executed in the Jido environment in Toulouse. A - B. Jido reaching for box and grasping. C – D. Jido puts box on red table mat.

C. *Experiment 3 – Cross platform generalization*

The Shared Plan Manager creates permanent definitions of these new actions, which can then be transferred via the SVN system for use on other robots at other sites. We could thus test the definition of `Cover X with Y` that was learned on the iCub in Lyon on the Jido robot in Toulouse.

Via the RAD Supervisor, the human asked Jido cover the red table-mat with the box (see Figure 3). The Supervisor retrieved the composite action definition, communicated the corresponding motor primitives corresponding to `grasp X` and `release X on Y` to Jido. Jido was thus able to produce the `cover X with Y` action, based on learning that had occurred on a morphologically distinct robot. Thus, despite this morphological difference, because of the abstraction at both perceptual and execution levels, action knowledge acquired on one platform can be exploited on another.

D. *Experiment 4 - Agency assignment with Kinect*

In behavior that involves object manipulation, the human hand has a special status as an agent. Indeed it has been shown that infants may prefer to assign agency to well known agents however they also rely on naïve physics and assign agency to objects that are moving on their own and in specific ways [4, 39]. In order to achieve accurate hand tracking we demonstrate here how the Kinect motion tracker can provide this capability. A module has been developed using the Kinect device in combination with OpenNI drivers[1] in order to track the user hands and add them to the Egosphere as standard objects. Since this module is on the platform specific side of the Egosphere, then no change is required to use its information. We achieved the same result using our standard vision system and visual markers on the human hand; however the approach with the Kinect is much more natural and robust. In the experiment the user was teaching system how to recognize *cover* and *uncover* and the system recognized these actions, and which hand performed them so it could describe it in the following way: "I detected that the *human hand* covered the *toy* with the *box*".

E. *Experiment 5 – Goal Directed Action Imitation*

This experiment, illustrated in detail in Figure 4, brings all of the functionality together. To arrive at this point, the robot should be able to both recognize and execute a set of actions. Here we demonstrate this with the `cover the toy with the box` action. This is illustrated briefly in Figure 2C and 2D. Figure 2C illustrates the human user showing the action to the robot. Figure 2D illustrates the robot now performing the recognized action. Full detail of

[1] Kinect is a hardware product by Microsoft (http://www.xbox.com/en-US/kinect). OpenNI.org release open source drivers for the Kinect device (http://openni.org/).

the experiment is provided in Figure 4. A video demonstrating this experiment is attached with the paper.

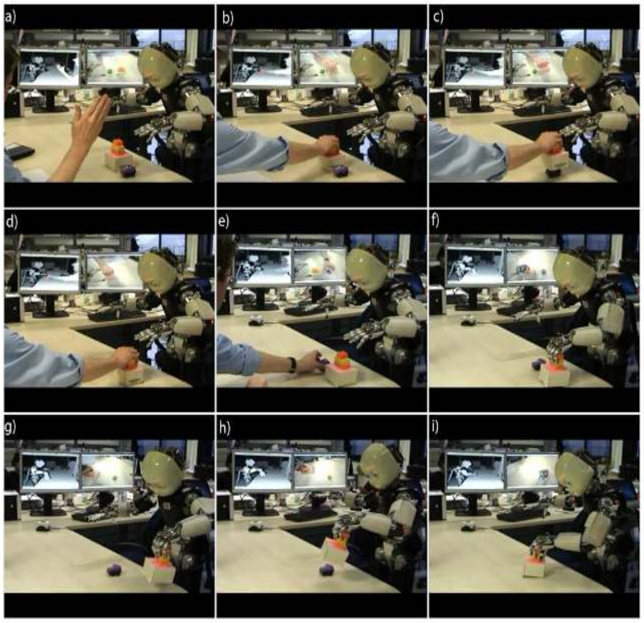

Figure 4: Experiment 5. Imitation. A. Calibration of hand recognition with Kinect. B-D. Human covers toy with box. E. Human repositions objects. F. Robot grasps box. G-I. Robot covers toy with box, completing the imitation.

VI. Discussion

Many of the mirroring skills demonstrated in the literature [40, 41] use the perceived motor state of the agent (i.e. its kinematic evolution over the action) to both recognize and execute actions. This has been combined with goal-based representations [10]. Our system is based on the fact that each action can be recognized by its perceptual consequences in changes in the world state (object states) and then performed by executing the associated motor commands. Those motor commands are not robot specific, but the primitives they call are, which implicitly solves the correspondence problem described in [8, 42]. Although we cannot argue that our system can cope with the same range of actions as a “trajectory based” systems, it is complimentary with such systems, and can be used at a higher level, for actions involving multiple arguments and symbolic goal achievement more than precise motor imitation. Indeed, this approach also emphasis the equifinal means of an action since the user can demonstrate an action and then the robot will achieve the same result with completely different trajectories.

Aspects of this work can thus be considered in the context of learning by imitation or demonstration, which is a major area of research in robot cognition today [8, 10, 40-42]. Our novel contributions to this domain include (1) the encoding of action in terms of perceptual state changes and composed motor primitives that can achieve these state changes, in a manner that allows the robot to learn new actions as perception – execution pairs, and then use this knowledge to perceive and imitate. (2) These actions can take several arguments, e.g. `AGENT put the OBJECT on the RECIPIENT,` which allows for the generalization of learned actions to entirely new contexts, with new objects and agents. This yields the equifinal component of action where the same goal can be achieved by different means. (3) We use spoken language interaction and visual perception to provide learning input to the system. In our long term research program, this provides that basis for learning to perform cooperative shared tasks purely through observation.

In our system actions are encoded using the effect they produce on the state of the world, the latter being abstracted in terms of unspecific quantities like relative position and orientation of objects and their visibility. The particular type of encoding we adopt for actions is therefore completely independent of the robot platforms, and can therefore be transferred between robots with different embodiments or perceptual systems. In previous work we showed how motor skills could be transferred between robots; this paper extends this work to action recognition and mirroring.

Our approach to action representation is consistent with and inspired by the 'teleological framework' [43, 44] that represents actions by relating three relevant aspects of reality (action, goal-state, and situational constraints) through the inferential 'principle of rational action', which assumes that: (a) the basic function of actions is to bring about future goal states; and that (b) agents will always perform the most efficient means action available to them within the constraints of the given situation. This approach is complimentary to existing approaches that take the “means” (e.g; aspects of demonstrated trajectories) into account [29, 36, 45]. Future research should consider how to combine these approaches.

VII. Acknowledgment

This research was supported by the European Commission under the Robotics and Cognitive Systems, ICT Project CHRIS (FP7-215805).

VIII. References

1. Lallée, S., et al. *Towards a Platform-Independent Cooperative Human-Robot Interaction System: I. Perception*. in *IROS*. 2010. Taipei.
2. Hommel, B., et al., *The theory of event coding (TEC): A framework for perception and action planning*. Behavioral and Brain Sciences, 2001. **24**(05): p. 849-878.
3. Rizzolatti, G. and L. Craighero, *The mirror-neuron system.* Annu. Rev. Neurosci., 2004. **27**: p. 169-192.
4. Király, I., et al., *The early origins of goal attribution in infancy.* Consciousness and Cognition, 2003. **12**(4): p. 752-769.
5. Dominey, P. and F. Warneken, *The basis of shared intentions in human and robot cognition.* New Ideas in Psychology, 2009: p. (in press).
6. Lallée, S., F. Warneken, and P. Dominey. *Learning to collaborate by observation*. in *Epirob*. 2009. Venice.

7. Csibra, G., et al., *Goal attribution without agency cues: the perception of [] pure reason'in infancy.* Cognition, 1999. **72**(3): p. 237-267.

8. Alissandrakis, A., C.L. Nehaniv, and K. Dautenhahn, *Imitation with ALICE: Learning to imitate corresponding actions across dissimilar embodiments.* Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2002. **32**(4): p. 482-496.

9. Argall, B., et al., *A survey of robot learning from demonstration.* Robotics and Autonomous Systems, 2009. **57**(5): p. 469-483.

10.Calinon, S., F. Guenter, and A. Billard. *Goal-directed imitation in a humanoid robot*. in *ICRA*. 2005. Barcelona: IEEE.

11.Demiris, Y. and M. Johnson, *Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning.* Connection Science, 2003. **15**(4): p. 231-243.

12.Dillmann, R., *Teaching and learning of robot tasks via observation of human performance.* Robotics and Autonomous Systems, 2004. **47**(2-3): p. 109-116.

13.Woodward, A.L., *Infants selectively encode the goal object of an actor's reach.* Cognition, 1998. **69**(1): p. 1-34.

14.Woodward, A.L., *Infants' ability to distinguish between purposeful and non-purposeful behaviors.* Infant Behavior and Development, 1999. **22**(2): p. 145-160.

15.Woodward, A.L., J.A. Sommerville, and J.J. Guajardo, *How infants make sense of intentional action.* Intentions and intentionality: Foundations of social cognition, 2001: p. 149–169.

16.Kamewari, K., et al., *Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion.* Cognitive Development, 2005. **20**(2): p. 303-320.

17.Mandler, J., ed. *Preverbal representation and language*. Language and space. 1996, MIT Press. 365–384.

18.Talmy, L., *Force dynamics in language and cognition.* Cognitive science, 1988. **12**(1): p. 49-100.

19.Kotovsky, L. and R. Baillargeon, *The development of calibration-based reasoning about collision events in young infants.* Cognition, 1998. **67**(3): p. 311-351.

20.Metta, G., et al. *The iCub humanoid robot: an open platform for research in embodied cognition*. in *PerMIS: Performance Metrics for Intelligent Systems Workshop*. 2008. Washington DC, USA.

21.Baillargeon Elizabeth, S., _Object permanence in five-month-old infants* 1._ Cognition, 1985. **20**(3): p. 191-208.

22.Mandler, J.M., *How to build a baby: II. Conceptual primitives.* PSYCHOLOGICAL REVIEW-NEW YORK-, 1992. **99**: p. 587-587.

23.Roy, D., *Semiotic schemas: A framework for grounding language in action and perception.* Artificial Intelligence, 2005. **167**(1-2): p. 170-205.

24.Siskind, J.M. *Visual event perception*. in *NEC Research Symposium*. 1998.

25.Spelke, E.S., et al., *Origins of knowledge.* PSYCHOLOGICAL REVIEW-NEW YORK-, 1992. **99**: p. 605-605.

26.Dominey, P. and J. Boucher, *Learning to talk about events from narrated video in a construction grammar framework.* Artificial Intelligence, 2005. **167**(1-2): p. 31-61.

27.Dominey, P. and J. Boucher, *Developmental stages of perception and language acquisition in a perceptually grounded robot.* Cognitive Systems Research, 2005. **6**: p. 243-259.

28.Mussa-Ivaldi, F.A., S.F. Giszter, and E. Bizzi, *Linear combinations of primitives in vertebrate motor control.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(16): p. 7534.

29.Mataric, M.J., et al. *Behavior-based primitives for articulated control*. in *Fifth international conference on simulation of adaptive behavior on From animals to animats 5*. 1998.

30.Williamson, M.M. *Postural primitives: Interactive behavior for a humanoid robot arm*. in *Fourth international conference on simulation of adaptive behavior on From animals to animats 4*. 1996.

31.Thomas, U., et al. *Error-tolerant execution of complex robot tasks based on skill primitives*. in *ICRA*. 2003. Taipei: IEEE.

32.Morrow, J.D. and P. Khosla. *Manipulation task primitives for composing robot skills*. in *ICRA*. 2002. Albuquerque: IEEE.

33.Sentis, L. and O. Khatib, *Synthesis of whole-body behaviors through hierarchical control of behavioral primitives.* International Journal of Humanoid Robotics, 2005. **2**(4): p. 505-518.

34.Firby, R.J. *Building symbolic primitives with continuous control routines*. in *First international conference on Artificial intelligence planning systems* 1992.

35.Paine, R.W. and J. Tani, *Motor primitive and sequence self-organization in a hierarchical recurrent neural network.* Neural Networks, 2004. **17**(8-9): p. 1291-1309.

36.Mussa-Ivaldi, F. and E. Bizzi, *Motor learning through the combination of primitives.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2000. **355**(1404): p. 1755.

37.Lallée, S., et al., *Linking language with embodied teleological representations of action for humanoid cognition.* Frontiers in Neurobotics, 2010.

38.Sutton, S., et al. *Universal speech tools: The CSLU toolkit*. in *Fifth International Conference on Spoken Language Processing*. 1998.

39.Song, H., R. Baillargeon, and C. Fisher, *Can infants attribute to an agent a disposition to perform a particular action?* Cognition, 2005. **98**(2): p. B45-B55.

40.Johnson, M. and Y. Demiris. *Hierarchies of coupled inverse and forward models for abstraction in robot action planning, recognition and imitation*. in *AISB*. 2005.

41.Metta, G., et al., *Understanding mirror neurons: a bio-robotic approach.* Interaction studies, 2006. **7**(2): p. 197-232.

42.Nehaniv, C.L. and K. Dautenhahn, *2 The Correspondence Problem.* Imitation in animals and artifacts, 2002: p. 41.

43.Gergely, G. and G. Csibra, *Teleological reasoning in infancy: the na ve theory of rational action.* Trends in Cognitive Sciences, 2003. **7**(7): p. 287-292.

44.Gergely, G., *What should a robot learn from an infant? Mechanisms of action interpretation and observational learning in infancy.* Connection Science, 2003. **15**(4): p. 191-209.

45.Pattacini, U., et al. *An Experimental Evaluation of a Novel Minimum-Jerk Cartesian Controller for Humanoid Robots*. in *IROS*. 2010. Taipei.