ROBotic Open-architecture Technology for Cognition, Understanding and Behavior

Project no. 004370

RobotCub

Development of a cognitive humanoid cub

Instrument: Integrated Project
Thematic Priority: IST – Cognitive Systems

4.1 Results of experiments on affordant behaviors

Due Date: M65
Submission date: M65

Start date of project: **01/09/2004** Duration: **60+5 months**

Organisation name of lead contractor for this deliverable: IST

Responsible Person: Prof. Alexandre Bernardino

Revision: **3.0**

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Table of Contents

Revision History 3

1 Introduction 4
- 1.1 Object's Affordances and the Developmental pathway 4
- 1.2 Organization of the Document 5

2 Results from neural sciences 6

2.1. Monkey experiments: influence of visual information 6
- 2.1.1. Introduction 6
- 2.1.2. Aim 7
- 2.1.3. Experimental device and paradigm 7
- 2.1.4. Results 8
- 2.1.5. Conclusion 8

2.2 Human experiments: a transcranial magnetic stimulation study on grasping a predictable moving object 9
- 2.2.1 Introduction 9
- 2.2.2. Experimental device and paradigm 9
- 2.2.3. Results 11

3 Results from psychological sciences 11
- 3.1 Influence of object form in hand orientation 11
- 3.1 Fitting objects into apertures 12
- 3.2 Piling objects 13

4 Results from experimental robotics 13
- 4.1 Self-exploration, learning through observation and imitation 13
- 4.2 Experiments at IST 15
  - 4.2.1 Affordance model 15
  - 4.2.2 From affordances to imitation 18
  - 4.2.3 Case of study: Learning good grasping points 18
  - 4.2.4 Labelling of affordance knowledge (word association) 20
- 4.3 Experiments at UGDIST 21
  - 4.3.1 Deriving probability distributions over action primitives 23

5 Conclusions 26

6 References 26

Revision History

Version 1.0 (M18)
Original Draft. First models and experiments.

Version 2.0 (M30)
Update. A statistical formulation of affordance learning and utilization, as well as and new robotics experiments.

Version 3.0 (This)
Update. Final experiments and results.

1 Introduction

The RobotCub project addresses the implementation of a humanoid robot's manipulative skills through learning, imitation and social communication. One key issue in this process is the ability to acquire and exploit knowledge about objects properties and their usage to achieve goals (object's affordances). The document reports results of experiments done within the RobotCub consortium, conducted at UNIFE, UNIUP, UGDIST and IST, aiming at the definition and implementation of the cognitive skills required for the acquisition/exploitation of object's affordances. We present results on the acquisition of objects affordances from studies both in neuroscience, developmental psychology and experimental robotics.

1.1 Object's Affordances and the Developmental pathway

The term affordance was originally used by James J. Gibson [1] to refer to all "action possibilities" on a certain object, with reference to the actor's capabilities. Thus, a chair is only "sit-able" for a perceiver of a certain height. However, whether an affordance is exploited by a perceiver or not has to do with the goals, values and interests of this perceiver.

Humans learn to exploit object's affordances through their entire lifespan but not all are learnt autonomously. A large set is conveyed by social means either by communication or by observing others actions. Due to the complexity of the human developmental process, it is difficult to separate the importance of learning by exploration and learning from others. Furthermore, learning from others may sometimes just be a question of highlighting a certain affordance. Notwithstanding, we split the problem in two fundamental means of acquisition of object's affordances: by self-exploration (autonomous learning) and by observation (learning from examples). On a developmental perspective, it is natural to consider that self-exploration precedes the observation stage, though they are not simply sequential stages. Learning by observation requires some minimal capabilities, such as object and action recognition, in order to infer other agents' actions on objects, which are capabilities acquired by previous self-interaction with the environment. After this initial stage, both modes coexist and reinforce each other.

The acquisition of object affordances also depends, developmentally, on the existence of some minimal perceptual and motor capabilities. It is essential to be able to individuate objects in the environment and execute directed prospective motor actions over objects. Much of the work in WP3 (sensorimotor coordination) focuses on the development of capabilities for controlling own actions which constitutes an important part of the primitives for the acquisition of object affordances. After the system has acquired the capability to coordinate its own actions, it can start interacting with objects and understanding its interface – how to grab an object, what are the effects of certain applied actions. Then, the system may start recognizing and interpreting other agents interacting with similar objects, learning other object's affordances and interpreting activities. These capabilities have important relationship with the development of imitation and gesture communication (WP5 and WP6).

1.2 Organization of the Document

The work presented in this deliverable contributes to the specific objective S0-3a, i.e, to model, implement and understand to a certain degree, the cognitive aspects underlying the ability of learning and exploiting object affordances in order to correctly grasp and manipulate objects on the basis of their use.

The problem is addressed from three complementary perspectives. In section 2 we present results from electrophysiological experiments conducted at UNIFE both on humans and primates. The study based on single neurons recording in the monkey concerns the influence that visual information of the grasping hand has on the activity of motor neurons in pre-motor and motor cortices. The study using transcranial magnetic stimulation in humans investigates the time-course of corticospinal excitability during the grasping of a predictable moving object (during both go and no-go conditions), and during the observation of the same task.

In section 3 we present results on developmental psychology conducted at UNIUP concerning infants' ability to fit blocks into apertures, pile blocks on the top of each other, and orient grasping to object orientation.

In section 4 we present experiments in robotic setups, divided in two parts.

The first part describes a Bayesian model for the acquisition of objects' affordances by self-exploration and observation, developed at IST. After learned, the model can be used for several tasks, from gesture recognition to planning and imitation. We present results on the self-exploratory learning capabilities of the system, on playing imitation games with users, on learning grasping points and on labelling affordance knowledge. The model presents a clear developmental link between the lower level sensorimotor learning and the imitation and communication higher levels.

The second part describes self-experimenting results conducted at UGDIST. It describes experiments that illustrate the autonomous acquisition of object models from manipulation, exploitation of object's properties for grasping, determination of successful grasps and the exploitation of the affordances Bayesian model for deriving probability distributions from over action primitives.

2 Results from neural sciences

2.1. Monkey experiments: influence of visual information

2.1.1. Introduction

It is well known that the frontal cortex is strongly involved in action programming and motor control. In addition to the primary motor cortex (area F1) there are three pairs of areas: F3 (caudal, SMA proper) and F6 (rostral, pre-SMA) lay on the mesial wall of the frontal lobe; F2 (caudal) and F7 (rostral) form the dorsal premotor cortex and F4 (caudal) and F5 (rostral) form the ventral premotor cortex. Particularly interesting are the ventral premotor areas because of the strong visual input they receive from the inferior parietal lobule. These inputs subserve a series of visuomotor transformations for reaching (area F4 [12]) and grasping (area F5 [18,16]). In addition, area F5 contains neurons forming an observation/execution matching system, which maps observed actions on the observer's internal motor representations (**mirror neurons**). Electrical stimulation studies revealed that area F5 contains extensively overlapping representations of hand and mouth movements [18,14]. Single neurons studies have shown that most F5 neurons code specific actions, rather than the single movements that form them [18]. It has been therefore proposed that, in area F5, a vocabulary of goals more than a set of individual movements, is stored. Several F5 neurons, in addition to their motor properties, respond also to visual stimuli. According to their visual responses, two classes of visuomotor neurons can be distinguished within area F5: canonical neurons and mirror neurons [17]. Canonical neurons respond to visual presentation of three-dimensional objects [16]. About one quarter of F5 neurons show object-related visual responses, which are, in the majority of cases, selective for objects of certain size, shape and orientation and congruent with the motor specificity of these neurons. They are thought to take part in a sensorimotor transformation process dedicated to select the goal-directed action, which most properly fits to the particular physical characteristics of the to-be-grasped object.

The mirror neurons form the second class of visuomotor neurons of area F5. This name was coined because of their property to "reflect" with their visual response an action executed by another individual, if the seen action is similar to that motorically coded by them [11, 13, 20]. In contrast to the canonical neurons, mirror neurons do not respond to the mere presentation of objects. Thus, the vision of a real action, performed by a biological agent (the experimenter or another monkey) is essential for their activation. A mimed action, not interacting with an object, or an action executed with a tool (e.g. pliers) is ineffective in triggering most of F5 mirror neurons. Almost all mirror neurons show a certain degree of congruence between the effective observed and executed action. This congruence is very strict in about one third of F5 mirror neurons. Very recently, it has been reported that a fraction of mirror neurons, in addition to their visual response, become also active when the monkey listens to an action-related sound (e.g. breaking of a peanut) [15]. It is tempting therefore to conclude that mirror neurons may form a multimodal representation of goal directed actions, possibly involved in action recognition. The recent finding that mirror neurons become also active when the effective observed action is partially hidden to the monkey [21], suggests that they may represent actions in a rather abstract and cognitive way.

2.1.2. Aim

The goal of monkey experiments was to investigate the nature of the visuomotor coupling at the basis of the “mirror” response. Our hypothesis was that mirror discharge could be initially generated by the observation of one’s own acting effector, seen from different perspectives, performing repetitively the same action. We assumed that these different visual information could be associated by the brain as “common signals”, having in common the same motor goal. Following this learning phase, the system could become therefore capable to extract motor invariance also during observation of actions made by others. Although the learning process described above should mainly occur during development, we postulated that also in adult animals some vestigial residuals of this visuomotor coupling could have resisted in F5 motor neurons (generally considered as devoid of any visual property). To investigate this hypothesis, we programmed a series of single neuron recordings in monkey premotor area F5 while the animal was executing a grasping movement with normal and manipulated visual information (e.g.: complete dark, brief flash of light during different phases of the movement). As a control, primary motor cortex neurons (area F1) have been recorded too.

2.1.3. Experimental device and paradigm

To standardize the grasping movement, a specially designed apparatus has been used. It consists of a box that was mounted at reaching distance (30 cm) in front of the monkey, with little pieces of food hidden inside (Figure 1).

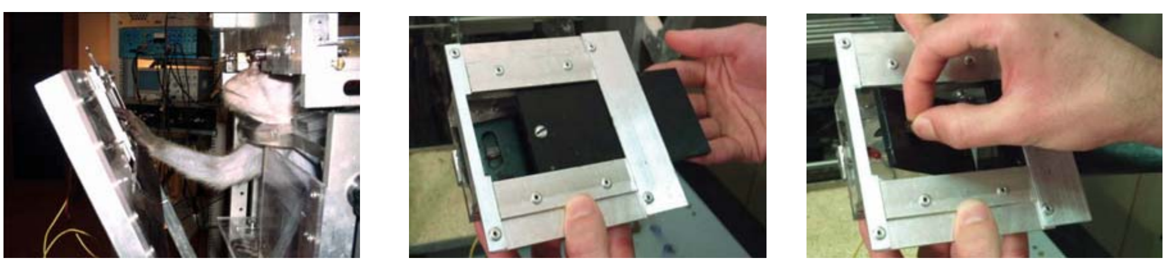


Figure 1 – The experimental apparatus

The box was covered by two doors. A more superficial one (see figure 2, left) whose opening at distance by the experimenter signaled to the monkey the beginning of the trial, and a second one (see figure 2, right), hosting a small plastic cube working as a handle. This plastic cube was translucent and back-illuminated from inside the box by a red LED in order to allow the monkey to fast reach it, also in the dark. The handle was buried inside a grove that forced the monkey to open the door by grasping the handle only by using a precision grip. When both thumb and index finger touched the handle, an electronic circuit (Schmitt’s trigger) gave to the acquisition system the synchronization signal. Neuronal activity was recorded during the two seconds following handle grasping, with one second of pre-trigger acquisition.

In order to test the experimental hypothesis, recorded neurons were submitted to four conditions:

a. grasping in full vision
b. grasping in dark with no hand visual feedback
c. grasping in dark with instantaneous visual feedback before contact
d. grasping in dark with instantaneous visual feedback at object contact

In the last two conditions a very brief (20 microseconds) xenon flash illuminated the scene at two different phases of the grasping action: during hand approaching (as triggered by a pyroelectric infrared sensor) (c) and at the moment of handle touch (d).

2.1.4. Results

We have collected preliminary results from the two hemispheres of the first monkey and now we are collecting data from a second monkey.
In the analysis of the recorded neurons we were particularly interested in neurons showing a reduction of their activity in the dark condition with respect to the light one. While in area F1 only about 15% of neurons satisfy this criterion, in area F5 about 57% of the recorded neurons reduced their activity when the grasping hand was not visible.

Moreover, neurons responses were subdivided into different epochs according to the phase of movement: Hand shaping epoch, from 250 ms before to the touch of the target handle with both thumb and index finger (precision grip); Touch/manipulation epoch, from handle grasping to 250 ms after (door opening). The statistical comparison between grasping with the hand fully visible (light condition) and grasping without hand vision (dark condition), in the two different epoch, showed that when the modulation is negative it mainly concerns the hand shaping epoch.

A further aspect of our analysis was concerned with the effect on neuronal discharge of a brief flash of light, which caused a sudden appearance of the acting hand. Although the dimension of our sample does not allow drawing a conclusive picture on neurons' behavior during flash conditions, these two conditions were included to control for the presence of phasic modulation of activity due to own hand vision. Few cells (about 10% of the modulated ones), showed this very specific phase-dependent modulation.

2.1.5. Conclusion

The results of monkey experiments presented in this deliverable are, in our view, of great interest. They firstly demonstrate that within a premotor area, involved in hand action programming and execution, there are motor neurons specifically modulated by the vision of monkey's own acting hand. The first important result achieved by these experiments is related to the direction of the modulation. In contrast with area F1, F5 motor neurons are negatively modulated by the absence of the visual hand. This reduction of the response could be, very likely, attributed to the lack of the hand-related visual input in this condition. The second result is that, when a negative modulation occurs, in general it involves the epoch preceding handle touching. If one consider that prediction is strongly embedded in feed-forward control systems, this anticipatory effect, specific for area F5, speaks in favor of a control role played by this area.

2.2 Human experiments: a transcranial magnetic stimulation study on grasping a predictable moving object

2.2.1 Introduction

Prompted by the discovery of monkey mirror neurons and stimulated by their possible involvement in high level cognitive functions, such as understanding others' behavior and interindividual communication, several functional brain imaging studies were performed to investigate whether or not a mirror-neuron system is also present in the human brain. Results showed that observation of an action recruits a consistent network of cortical areas, including the ventral premotor cortex (which extends posterior to the primary motor cortex), the inferior frontal gyrus, the inferior parietal lobule and the superior temporal cortex (for recent literature see Rizzolatti and Craighero [19]). However, brain imaging studies give us a static picture of the activated areas and do not enable us to conclude that the observer's motor system is dynamically (on-line) replicating the observed movements. Transcranial magnetic stimulation (TMS) can be used to measure the corticospinal (CS) excitability with a relatively high temporal resolution, and has been used extensively to address this issue.

We are currently investigating by TMS the time course of corticospinal excitability during the execution (during both go and no-go conditions) and the observation of a catching task, in order to understand, firstly, how the motor system behaves during such a phasic motor action and, secondly, if and how it is involved during observation of the same action.

2.2.2. Experimental device and paradigm

Interception in humans is a complex visuo-motor task that requires in few hundreds of milliseconds to detect and process visual motion information, to estimate future position of object in space and time, to transform visual information into an appropriate motor action and to trigger this action in advance to compensate for physiological and mechanical delays. Despite this complexity, humans demonstrate rather good performance in interceptive actions. To investigate this ability and understand the characteristics of the underlying visuo-motor transformation, we estimated individuals' corticospinal facilitation by means of TMS at different time intervals during the phase immediately preceding an interceptive task of a falling object, in three different experimental conditions: when participants were required to catch a falling object, when they were asked to observe an agent catching it, and when they had to voluntarily refrain from catching it.

Subjects are required to catch a cylinder which is moving along a vertical bar and it is automatically released by the computer (Figure 2). A beep is presented at the instant of releasing. The position of the subject and of that of the bar is such to allow the subject to catch the cylinder by simply closing her hand at the appropriate time, without moving her arm or body. Single pulse TMS is delivered over the *first dorsal interosseus* cortical motor representation at four different intervals with respect to the releasing of the cylinder: 200 ms before, 0 ms, 100 after, 200 ms after (Figure 3).

Three are the different experimental conditions:

- Real Interception: the subject has to execute the task.
- NoGo Interception: the subject has to observe the falling cylinder and to refrain herself in catching it.
- Observation Interception: the subject has to observe the experimenter executing the task.

Motor evoked potentials (MEPs) are recorded from the *first dorsal interosseus* muscle, which is a hand intrinsic muscle involved in the movement required to the subjects for catching the cylinder.

Figure 2: Experimental set-up. A tube was sliding along a vertical staff, passing between the participants thumb and fingers. The participants had to intercept (Execution condition) the bar, to observe a real actor intercepting the bar (Observation condition) or just looking at the falling of the tube (No-go condition).

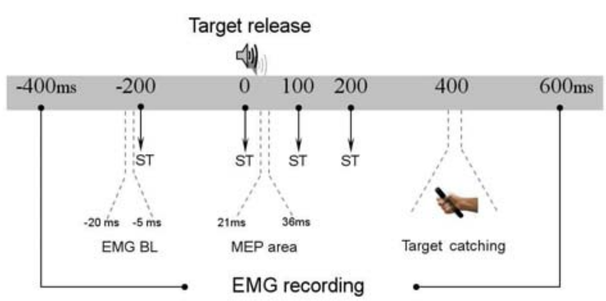


Figure 3: Schematic representation of the time course of experimental events. Arrows indicate the four stimulation times (ST) during Observation, Execution and No-go conditions. The baseline (BL) of the EMG activity was calculated before each ST and the area of the Motor Evoked Potential (MEP) after each ST. They are presented only once on the figures for clarity.

Our aim is to compare corticospinal excitability during the different phases of the task, when really executed, in order to have indications about the timing strategies for prospective control of grasping. Moreover, the other experimental conditions allow us to compare corticospinal excitability during observation of the same task. In our hypothesis, based on the literature on *mirror* neurons, the motor system is critically involved in both conditions. Finally, the no-go condition is compared to the execution condition in order to see if the involvement of the motor system is mandatory whenever an action-involving situation is presented.

2.2.3. Results

Results indicate that the involvement of the motor system during the preparatory phase of an interceptive task differs according to the tested experimental conditions [31]. In particular, the execution and the voluntary inhibition of this action determine a completely different modulation of the motor system (Figure 4). During Execution condition corticospinal activity is almost always increased with respect to the baseline, a part from the period immediately after target releasing. This decreasing in excitability is considered, on the basis of previous studies, an indication of the readiness to react. During No-go condition corticospinal activity doesn't differ from baseline a part from a significant inhibition at the time approximately corresponding to the mean EMG onset during actual execution, indicating that participants were effortfully paying attention to the falling target. Finally, the mere observation of an agent preparing to execute an interceptive task, when the exact instant of action execution is perfectly known by the observer, is not sufficient to elicit a corticospinal modulation. Consequently, present results do not support the hypothesis that the motor system involvement during action observation is functionally equivalent to motor preparation, furthermore, they suggest that motor representation activation is present in the observer only during the perception of the actual execution of another individual's action.

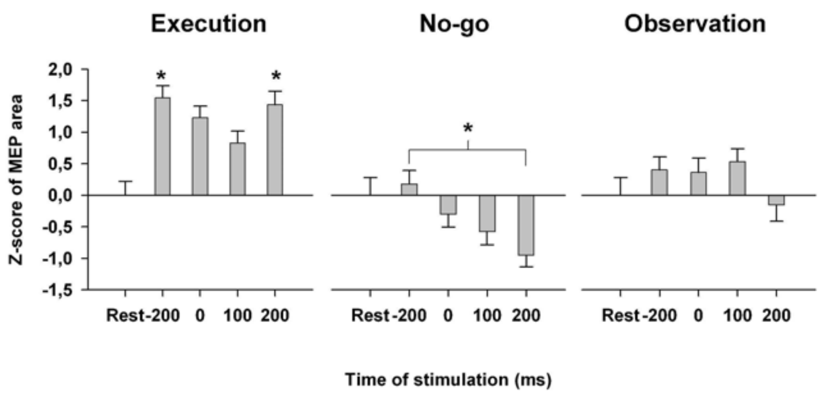


Figure 4: Time course relative to bar release (t = 0 ms) of Z-score of averaged FDI MEP area during execution (left), inhibition (middle) and observation (right) of target interception. Vertical bars represent standard error. Isolated asterisks denote statistically significant difference ($p < 0.05$) relative to respective baseline.

3 Results from psychological sciences

The Uppsala group has primarily worked on three problems. First, we have studied how infants' perception of object form determines reaching behaviour. Secondly, we have studied how infants' understanding of the relationships between objects and apertures, thirdly how their understanding of objects and gravity makes them able to construct towers.

3.1 Influence of object form in hand orientation

From the age when infants start to reach for objects they have been found to adjust the orientation of the hand to the orientation of an elongated object reached for [25,23,24]). Von Hofsten and Johansson (2005) [24] found that, when reaching for a rotating rod, infants prepare the grasping of the object by aligning the hand to a future orientation of the rod. Adjusting the hand to the size of a target is less crucial. Instead of adjusting the opening of the hand precisely to the size of the object, infants tend to open the hand more fully. Von Hofsten and Rönnqvist (1988) [22] found that 9 and 13 month-old infants, but not 5-month-olds, adjusted the opening of the hand to the size of the object reached for. They also monitored the timing of the grasps and found that the infants started to close the hand before the object was encountered. For infants of 9 months and younger the hand first moved to the vicinity of the target and then started to close around it. For the 13 month-olds, however, the grasping action typically started during the approach, well before touch. In other words, at this age grasping started to become integrated with the reach to become one continuous reach-and-grasp act.

3.1 Fitting objects into apertures

Handling objects reveal their different affordances and knowledge about affordances improves the handling of them. The close connection between vision and manipulation makes it possible to learn about object affordances by viewing other people manipulating objects. This is especially relevant when learning about the functions of tools. The development of skills in reaching and manipulation are closely related to the development of such cognitive skills as mental rotation and means-end relationships. When manipulating objects, the subject needs to imagine the goal state of the manipulation and the procedures of how to get there. Von Hofsten & Örnkloo (2005) [26] studied how infants develop their ability to insert blocks into apertures. The task was to insert elongated objects with various cross-sections (circular, square, rectangular, elliptic, and triangular) into apertures in which they fitted snugly. All objects had the same length and the difficulty was manipulated by using different cross sections of the objects. The cylinder fitted into the horizontal aperture as long as its longitudinal axis was vertical, while all the other objects also had to be turned in specific ways. The objects were both presented standing up and lying down. It was found that although infants younger than 18 months understood the task of inserting the blocks into the apertures and tried very hard, they had no idea of how to do it. They did not even rise up elongated blocks. They just put them on the aperture and tried to press them in. The 22-month-old children systematically rose up the horizontally placed objects when transporting them to the aperture and the 26-month-olds turned the objects before arriving at the aperture, in such a way that they approximately fit the aperture. This achievement is the end point of several important developments that includes motor competence, perception of the spatial relationship between the object and the aperture, mental rotation, anticipation of goal states, and an understanding of means-end relationships. The results indicate that a pure feedback strategy does not work for this task. The infants need to have an idea of how to reorient the objects to make them fit. Such an idea can only arise if the infants can mentally rotate the manipulated object into the fitting position. The ability to imagine objects at different positions and in different orientations greatly improves the child's action capabilities. It enables them to plan actions on objects more efficiently, to relate objects to each other, and plan actions involving more than one object.

These studies have been continued with two kinds of tasks. The children either had to choose which one of two objects fitted into a specific aperture or which one of two apertures a specific object would fit into. Children solve such problems significantly later than the one-object-one-aperture. We have found that it is not until the children are over 3 years old that they solve these problems in a consistent way. It seems to be the choice itself that is difficult. Choosing one solution requires that the alternative one is somehow inhibited. Children who have the ability to solve the object-aperture problem do not yet seem to be able to make such a choice. Similarly, 2-year-olds not yet choose the correct size of objects to be fitted into apertures. Given two objects, one that fits snugly into an aperture and one that is twice the size of it, they make random choices.

3.2 Piling objects

In collaboration with professor Rachel Keen at University of Massachusetts, we have started to study children's ability to construct towers out of blocks. The higher the tower, the more gently should the block be positioned on the lower block, in order to preserve the construction. We have found that 18-20-month-olds approach this problem in a variety of ways. Some children can build towers consisting of up to 5 blocks while others hardly manage a two-object tower. The successful children consistently approach the task in a more gently way than other children of the same age.

4 Results from experimental robotics

In this section we present results on the acquisition of object's affordances performed in robotic setups available within the consortium. First, we frame affordances in the iCub's developmental pathway, specially the links with sensory-motor coordination and with imitation.

Second, we describe the experiments conducted aiming at the acquisition of object's properties and affordances through manipulation and observation of others' actions. Some of the presented experiments are executed in robotic platforms not identical to the iCub, but this is not a limiting factor since the same principles are easily be adapted to the iCub system.

4.1 Self-exploration, learning through observation and imitation

As mentioned before, affordances develop on top of a minimal set of perceptual and motor capabilities. Much of this knowledge is provided by the sensory-motor coordination development (WP3).

While self-experimenting with objects, humans are able to learn a significant set of objects affordances. However, a purely autonomous approach search strategy would not allow, or make too long, the acquisition of more specialized utilizations of objects. Input provided by a teacher can reduce the learner's search space, speeding up the normal acquisition of object's affordances and introduce novel ones. In order to incorporate teacher's knowledge, it is required however, the existence of a representation layer common to both teacher and learner [3], e.g. it must be able to recognize some objects (or object's properties), actions and their effects on objects, and, at a certain stage, teacher's intentions and goals. Furthermore, teacher and learner must be compatible[1] such that, from observation data, the learner can match the state of the teacher to its own representation and be able to infer "hidden" information, like motor actions, gestures, goals, etc.

[1] A learner is compatible with a teacher when it has similar motor and perceptual core capabilities, though not being yet completely developed.

At this point, it is important to distinguish between "explicit" and "hidden" observation. When the learner is self-interacting with objects, several perceptual modalities are involved in the representation of the experience. For instance, when grasping an object, the agent has both motor information and visual information. However, when observing a teacher performing the same action, the agent only accesses visual information. Since there is evidence that recognition processes are based on motor information, the learner should infer motor information from visual. We thus consider visual information as "explicit" and motor information as "hidden", in the context of imitation. The same happens with goals and intentions. While self-experimenting, the actor knows its own goals, but when observing the teacher, they can only be inferred by "explicit" perceptual information (visual, auditory).

The "inference" procedure to match teacher's to learner's state may have several levels of complexity, from low-level relationships between muscular activation and visual perception, to abstract cause-effect relationships between actions/objects and goals/intentions. In principle, the richer is the learner's past experience, the more precise and abstract the inference can be. This assumption is consistent with generalist theories of imitation learning that consider that a person's imitation capabilities (including the interpretation of other's gestures and object interactions) depend on their past experience [4]. Given a certain level of base knowledge, learning by observation requires the transfer of teacher's state to observer's state and from two different sources of information: (i) current observations and (ii) prior knowledge. To blend these two sources of information we apply a Bayesian inference model. Bayesian methods are getting increasing attention in cognitive and developmental psychology due to their ability to represent cause-effect relations and incorporate prior knowledge with novel observations [8]. We have applied successfully this approach in the context of gesture recognition in [5].Once the teacher-learner bond has been established, we may consider two modes of learning: (i) belief revision and (ii) imitation learning. In some cases, the learner may already have enough prior knowledge to recognize/learn the object's affordance without any further training, for instance when objects action and goals are similar to ones experienced before (if an infant knows how to sit on a chair, it does not require too much effort to sit on a sofa). Therefore, the learner just needs to update the belief that a certain action on a certain object serves a determined purpose, given that the action, object and purpose are familiar. In other cases however, the observed situation is not similar to previously experienced ones and the learner may have to engage in imitation behaviour. In imitation learning, the observation drives the learner to autonomously interact with the object in the form suggested by the teacher, in order to learn more about objects' affordances as in the trial-and-error mode.

At this point, it is important to note that object affordance and the learning mechanisms described before place the robot in a development situation where imitation appears almost naturally. This is because, once the robot is able to recognize and learn from another agent, it has developed some of the basic building blocks required for imitation. In addition to this, affordances provide a link between sensory-motor coordination and higher level cognitive skills. By capturing the fundamental properties of the objects according to the agent perceptual and motor skills, they provide a way to perform prediction, recognition and planning which can be directly used in imitation.

4.2 Experiments at IST

The IST group has worked mainly in the development of a general model for affordance knowledge acquisition. This model is bootstrapped by self-experimentation. Once the model structure is learned, its parameters can be continuously updated based on the robot own experience and by observing other robots or humans. This model has been embedded in the developmental framework of [29]. We have conducted several experiments to validate the model and illustrate how affordances play an important role to link sensory-motor coordination (WP3) and imitation (WP5).

4.2.1 Affordance model

Building on previous work on Mirror neurons [5], we propose using Bayesian Networks (BN) to model the dependencies between robot actions, object characteristics and the resulting effects. Briefly, a BN [10] is described by a set of nodes that represent random variables, a set of directed arcs that encode conditional dependencies and a set of conditional probability distributions. A BN *encodes causality* since an arc from a node *X* to a node *Y* can be interpreted as *X* causes *Y*[2].

We assume that the robot has available, at the beginning of the affordance learning phase, the following skills[3]:

1- A motor action repertoire (*A*) – from its experience with the environment the agent has developed a repertoire of general motor actions, $\{m_j \in A\}$ (e.g. reach, grasp, throw), and associated visual measurements $\{v \in V\}$ (e.g. hand shape, and motion).
2- An object features repertoire (*F*). Based on observation of the world, the robot has developed a set of filters to detect basic object properties such as colour, shape, orientation, size, etc. As mentioned in Section 4.3, object manipulation facilitates the individuation of objects, $\{o_i \in O\}$, from the remaining environment, and the creation of perceptual models composed of object's properties, $\{f \in F\}$.
3- A set of resulting effects *(*E*)*. In most cases, the effects correspond to changes in the perceptual information. Some of them can also be acquired by observation of the world, but, as for the features, some self-experimentation may be needed

In order to learn the dependencies between the set of actions, objects and features, the robot interacts with the objects around it and measures the effects of its actions on different objects. This information is then used to estimate the Bayesian Network that maximizes the posterior probability (see Figure 5). This can be done using different learning algorithms [10].

Once the robot has learnt the structure of the network, i.e. the arcs between the nodes, it computes the parameters of the resulting conditional distributions. These parameters can be updated online as the robot performs more experiments. Also, they can be updated by observation of other agents. Further details about the learning algorithms can be found in [27].

[2] The ability to infer causality (instead of simply correlations) is still controversial. By using interventional data, we understand causality as statistical signatures obtained from experience.
[3] These skills have to be present although they may still need further development to reach their full operational state.

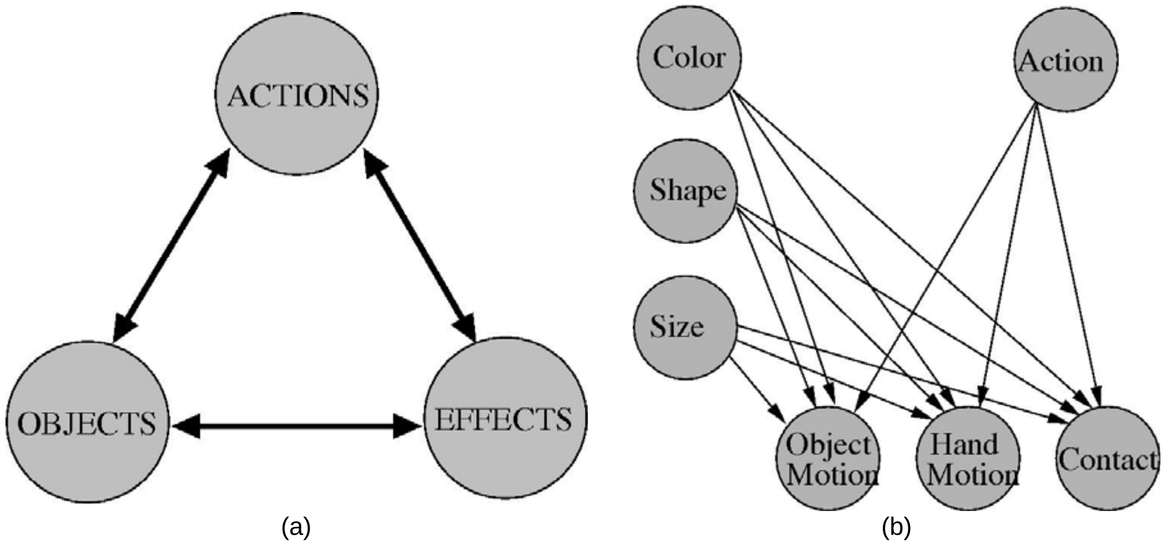


Figure 5: (a) General affordance scheme relating actions, objects (through their characteristics) and the resulting effects. (b) A particular BN encoding affordances.

The main characteristics of the model are:

1- Affordance learning through self-experience
2- Feature selection (or detection of irrelevant features)
3- Affordance learning through self-observation (restricted to the update of the probability distributions).
4- Usage of the model to perform prediction, recognition and planning. The use of the network is done based on probabilistic queries. These queries may take as input any combination of actions, objects and features and compute conditional distributions of one or more of the other variables. The following table summarizes some of the basic operations that can be performed with the network:

inputs	outputs	function
(O, A)	E	Predict Effect
(O, E)	A	Recognize action & Planning
(A, E)	O	Object recognition & selection

Table 1: Using affordances for prediction, recognition and planning.

Based on the previous model, we have performed several experiments with the robotic platform shown in Figure 6a. We used a playground scenario consisting of several objects with two shapes (box and ball), different sizes and colours. The robot was able to perform three different actions: grasp, tap and touch.

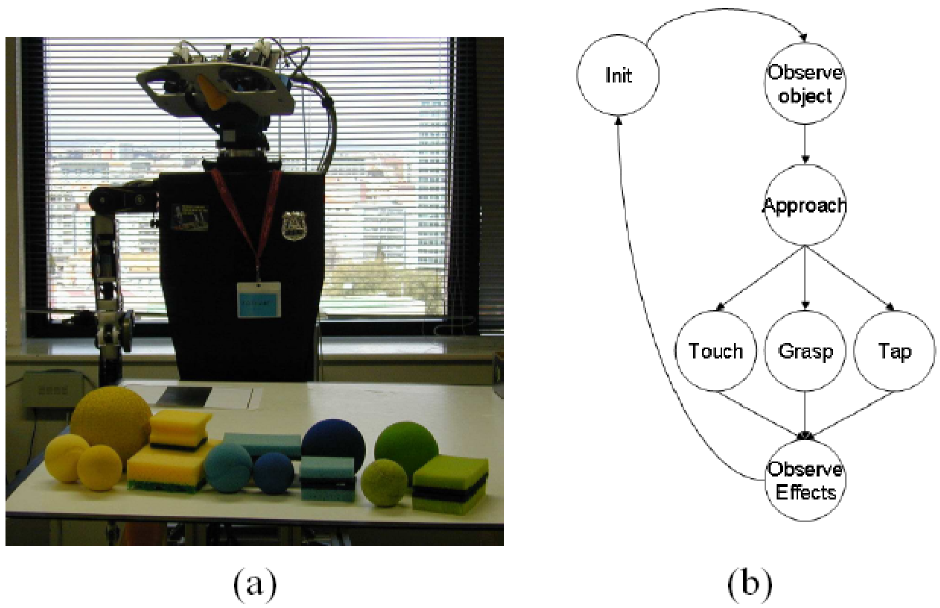


Figure 6: (a) The playground for the robot contains objects of several sizes, colours and shapes. (b) Experiments protocol. The object to interact is selected manually and the action is randomly selected. Object properties are recorded in the INIT to APPROACH transition when the hand is not occluding the object. The effects are recorded in the OBSERVE state. INIT moves the hand to a predefined position in open-loop.

The robot performed 300 experiments according to the protocol depicted in Figure 6b. The resulting affordance network is

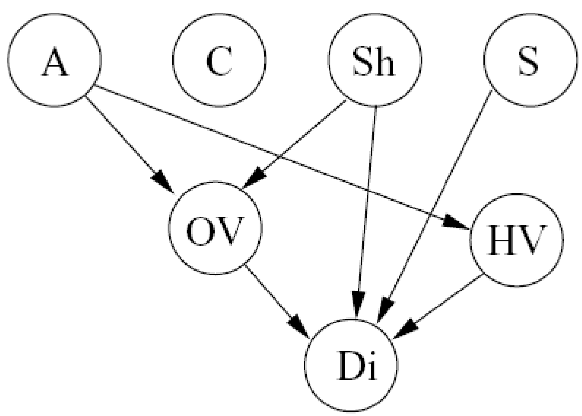


Figure 7: Learnt network. The variables represent A- Action, C- Object Colour, Sh- Object Shape, S- Object Size, OV – Object velocity profile, HV – Hand velocity profile, Di – Hand object distance profile.

The results show how the model is able to capture the basic object behaviour under different actions. For instance, colour is irrelevant in our setup. The shape has an effect on the object velocity (OV) and distance (Di) since tapping a ball or a box results in different effects (boxes do not roll). As expected, the hand velocity (HV) only depends on the selected action. The object hand distance (Di) also depends on the size since very big objects cannot be grasped by the robot. It is important to note that these relations are shaped by the experience of the robot and by its current skills. Another important property is that the detection of object features and effects is not perfect and the system has to cope with errors. In the same way, the same action on the same object does not always produce the same results. The probabilistic representation inherent to BN allows capturing and coping with this uncertainty. These results have already been submitted for publication and are currently under revision.

4.2.2 From affordances to imitation

Interestingly, the proposed affordance model also provides useful information to interact with other agents. When observing other agents, the learner must be able to match the observations to its own knowledge and interpret the behaviour of the teacher in terms of its past experience. In a certain observation experiment, the agent has direct access to "explicit" observations, i.e. object's visual features, f, visual perception of teacher's actions, v, and the perception of the goal effects, e, (in causal order). The transfer of knowledge from teacher to learner implies inferring the "hidden" non-observed information from the observed one and from prior knowledge. We propose a Bayesian formulation to this inference problem. The mentioned capabilities are directly provided by the affordance BN using the probabilistic queries presented in Table 1. In particular, one can recover:

1. Observation model for objects - the likelihood of observing certain object perceptual properties, provided that a particular object the object is present in the scene: $p(f \mid o_i)$. A problem in acquiring this knowledge is that objects can be observed in many different poses, so the model assumes that, while interacting with objects, the agent is able to learn its visual appearance from multiple poses.
2. Observation model for motor actions - the likelihood of the visual features elicited by each motor action: $p(v \mid m_j)$.
3. Observation model for goals – the likelihood of observing visual properties, given that a particular goal has been achieved $p(e \mid g_k)$.
4. The object's affordances model - the likelihood of a goal being achieved, given that a determined motor action is applied on a certain object: $p(g_k \mid m_j, o_i)$. This function codes the success on the utilization of a certain motor action over an object, for a certain purpose (or goal).

Priors for the occurrence of objects and motor actions, $p(o_i), p(m_j)$. Different models may exist for different situations and contexts.

The details of the developmental integration of affordances and the application to basic imitation games can be found in [27].

4.2.3 Case of study: Learning good grasping points

Graspable is a clear example of object affordances, and one that plays a very important role when interacting with objects. In addition to know if an object is graspable or not, it is important to know which part of the object actually allows the grasp action. For instance, the handle of a cup of tea or the narrow side of a pen are good grasping points of graspable objects. It is obvious that not all object parts are good grasping points and that this classification depends on the particular morphology of the robot, type of grasp and final objective of the task. Figure 8 illustrates graphically the model for a grasping point based on a set of visual descriptors for a cup of coffee.

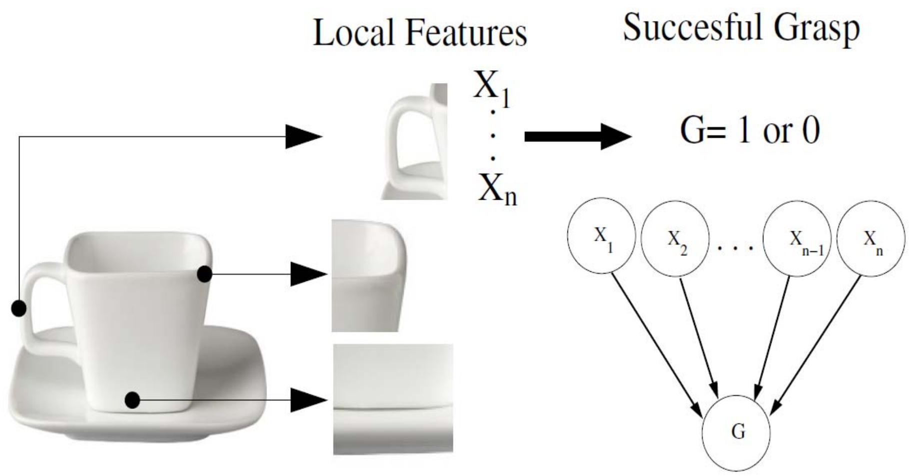


Figure 8: The probability of a successful grasp at a particular point depends on a set of features of the point.

In order to model grasping points from visual information, we developed a new algorithm [28] to estimate the probability of a successful grasp based on a set of local visual descriptors of the object extracted from a single camera and a set of labelled examples. This work draws inspiration from [33] where a simple logistic regression was used to classify grasping points based on a large set of examples. Our method is based on a non parametric kernel approximation of the probability at a point with features $\mathbf{x}_*$,

$$\bar{p}_* = \frac{\sum_{i=1}^{n} S_{*i} + \alpha_0}{\sum_{i=1}^{n} S_{*i} + \alpha_0 + \sum_{i=1}^{n} U_{*i} + \beta_0}$$

where

$$\begin{aligned} S_{*i} &= K(\mathbf{x}_*, \mathbf{x}_i) S_i \\ U_{*i} &= K(\mathbf{x}_*, \mathbf{x}_i) U_i \end{aligned}$$

are kernel accumulated counts for successful and failure grasps on the training database. These virtual counts are computed as a weighted sum of the success S_i and failures U_i at each point $\mathbf{x}_i$ of the database. The weights are provided by the kernel K(.,.) between the examples and the point of interest $\mathbf{x}_*$. Keeping counts on successes and failures allow us to use a Beta-Binomial model to model the full posterior probability of a grasping point

$$p(p_* \mid \mathbf{x}_*, \mathbf{X}_n, \mathbf{Y_n}) = \mathrm{Be}\left(p_*; \sum_{i=1}^{n} S_{*i} + \alpha_0, \sum_{i=1}^{n} U_{*i} + \beta_0\right)$$

In our case, where the robot learns by experimenting directly on the objects, this extra information allows for an active exploration strategy that can reduce greatly the amount of training required to discover good grasping points. For instance, one can use the variance of the potential grasping points to select the one that will provide more information, or could balance exploration and exploitation by combining variance and the probability of a successful grasp. Figure 9 shows some examples of the probability maps for different objects.

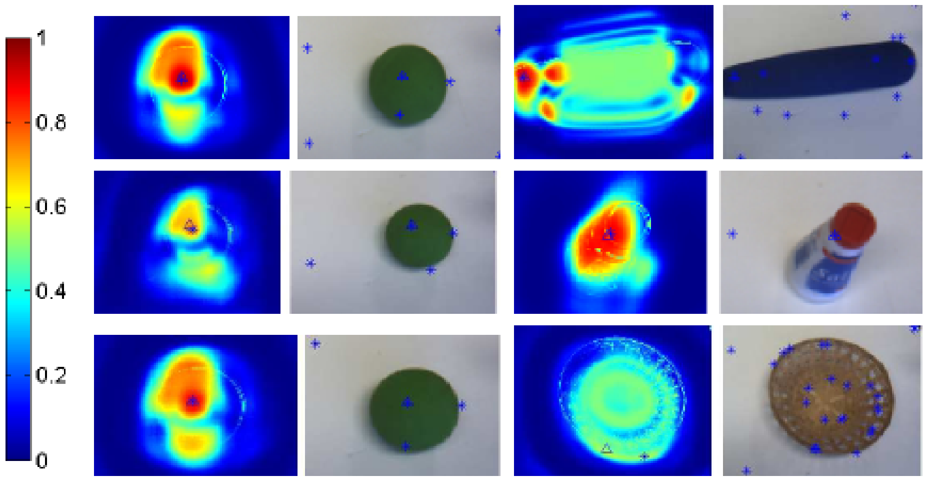


Figure 9: Probability maps for different objects. Three maps for the same object (left column) and three maps for more complex objects (right column).

The results of this work have been published in [28]. Results on active strategies are currently under review and we expect will be published on a journal during next year.

4.2.4 Labelling of affordance knowledge (word association)

The object affordances model described above captures relations between actions, objects and effects. These relations are basically common behaviours of the objects after an action. We have explored the possibility of using the affordance model to assign labels to this behaviours based on the patterns discovered by the network. The main idea is that these labels will correspond to the basic world properties represented in the network and will provide a common knowledge to interact with people or other robots abstracting from raw motor and visual information.

To provide the labels we used a speech recognition system (see Figure 10) and provided a verbal description based on a limited set of words. Most of the words directly correspond to values of the affordance node, for instance, colours such as red, blue or yellow; other words such as falling do not have a direct representation in the network but depend on a combination of the different nodes. There are also other words which do not have any node in the network (robot, it, is ...).

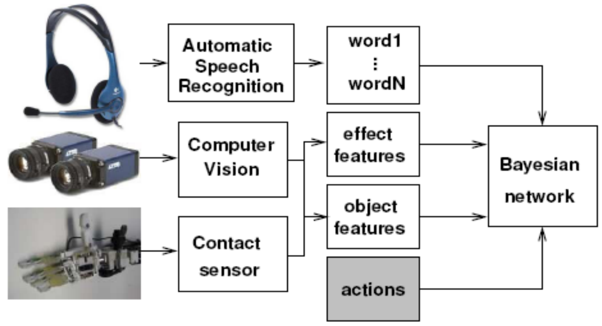


Figure 10: Experimental setup including the automatic speech recognition system.

Based on these verbal descriptions, we extended the affordance model to include the words and exploited co-occurrence to estimate the probability of the appearance of each word given the action, object features and effects. Figure 11 depicts a schematic view of the extended model.

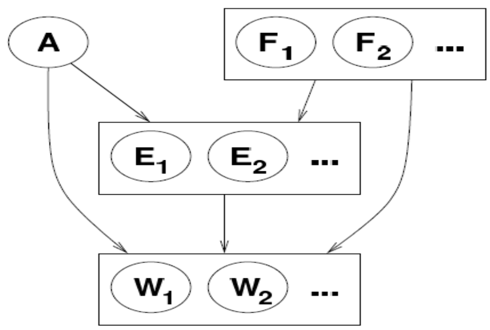


Figure 11: Extended network including the words as nodes.

The extended model is able to assign nouns and adjectives to the appropriate feature nodes and verbs to actions. These connections are usually one to one connections and have been demonstrated in the literature before. The ability of the model to discover more complex patterns and associate them to adverbs such as fast or slow that usually corresponds to effects is more interesting. The results of this work have been published in [29,30] and contain a full description of the model and the results.

4.3 Experiments at UGDIST

[This introduction is from v1.0 of this document, section 4.3.1 is new material]

On a developmental approach to humanoid robotics, a robot starts learning about if its own body, and successively about certain parts of the environment, including object properties and affordances. At UGDIST, experiments have been conducted following this approach. In particular, the experiments illustrate three important issues in acquiring object's related knowledge:

1. Grasping objects and modifying their pose through manipulation in order to obtain visual models of objects, from different perspectives.
2. Determining if a grasp on an object is successful or not, which is an essential feedback signal to learn the object's grasping affordances.

3. Adjusting hand posture and grasp plan according to object pose.

These topics are described in detail in papers [6] and [7], and are integrated in a developmental sequence, including sensorimotor coordination levels. Before being able to acquire object related knowledge, the system must be able to coordinate it own body and have a primitive ser of motor and perceptual skills, e.g. recognizing its own hand and reaching objects based on visual information. This is the subject of WP3 and is exploited here to provide an integrated view of the developmental pathway toward object perception.

The experiments have been performed on the robotic platform Babybot (see Figure 12). In the following lines we summarize the performed experiments addressing each of the mentioned topics.

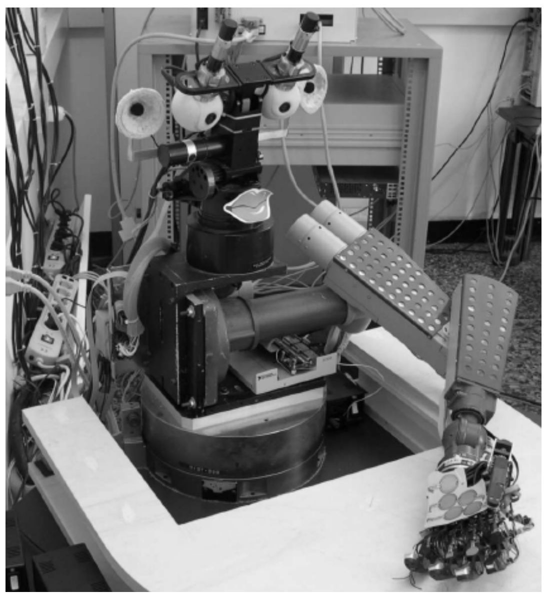


Figure 12: The robotic platform Babybot

Regarding the acquisition of object models, the robot starts by grasping an unknown object, either because a collaborative human has provided the object to the robot, or because the robot has managed to grasp it autonomously. The robot then starts the exploration of the object by bringing it close to the cameras in several different positions and orientations. During the exploration the robot keeps fixation on the object by tracking the hand with its internal kinematics model. At each position a few images are acquired and processed to train a part based model of the object in a probabilistic framework. After the exploration is completed, the robot has created a rather complete model of the object, allowing him to identify the same object in different positions and poses.

To determine if a grasping action is successful or not, the robot checks the weight of the object and its “consistence” in the hand (the shape of the fingers around the object). The intrinsic elasticity of the hand facilitates grasping, because fingers automatically adapt to the shape of the object, which allows the perception of the object by evaluating the hand shape. In case of failure another grasping trial is attempted (autonomous mode), or the robot waits for a new object to be placed in the palm (aided mode).

To maximize the success of grasping objects a disparity map of the segmented object is computed to determine its 3D orientation. Two different actions are then attempted: (i) if the principal axis is oriented horizontally the robot moves the hand above the object, otherwise, the hand approaches the object from the sides.

4.3.1 Deriving probability distributions over action primitives

Affordances are the perceived action possibilities offered by the environment and building robots that can perceive them promise to be an important stepping stone towards cognitive robots. A recent example of the successful application of the affordance based approach in the domain of dexterous manipulation is given by the work of Montesano et al. (IST). Their work builds on the idea of using Bayesian Networks (BNs) to formalize affordances. These Bayesian Networks are learned to encode relationships between objects, actions and observed effects to one another. In related work, Metta and colleagues (past work previous to RobotCub) described a desirable affordance representation to be probability distributions over action primitives given an object. The current work at IIT first aims at replicating the positive results obtained by IST in the domain of grasping actions, then, the problem how the obtained Bayesian Networks can be used to derive probability distributions over action primitives is addressed. It could be shown that in general Bayesian Networks are a suitable formalism for deriving affordances from experience. However, how well these affordances reflect the acquired experience proved to depend heavily on the topology of the network. The results suggest that the taken method generalizes to more complex tasks, but also that it needs to be complemented with mechanisms for inducing goal-oriented behaviour in robots.

The reader is referred to [32] (available from the RobotCub list of papers) for a complete description of the experiments. In the following a brief account of the main results is presented. Data for the experiments were collected with the iCub. A set of objects and primitive grasp types were used. In particular, the experimental protocol was as following:

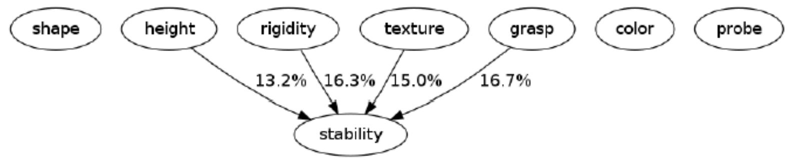
1 bringing the hand to the object
2 grasping the object with one of four different grasps
3 probing the object stability with one of two different moves
4 assessing the grasp stability
5 releasing the object
6 bringing the hand back to the starting position

Four actions, all considered primitives, were designed. In order to be better able to evaluate the affordance prediction capabilities of the resulting system, all four actions were pre-programmed and furthermore tailored to one specific reference object for which they should yield maximal stability. In addition to the grasp type, the second component making up an action that iCub applied to each object was a probing movement used to assess the stability of a given grasp/object combination. Two probing actions were implemented: probe encoding rotation of the forearm of approximately 150 degrees and “up and down” movement of the shoulder and elbow joint. As mentioned, the probing movement served to assess the stability of the grasp. Naturally, there are different possibilities for defining grasp stability. The most prominent measures for grasp stability are force and form closure and are drawn from an engineering perspective. Force closure and the related form closure are calculated by analyzing the forces exerted from the finger tips or the palm assuming certain friction coefficients and defining a stable grasp as one where arbitrary perturbations cannot alter the position of the grasped object. These measures, however, are unsuitable for our purposes. First, the geometric analysis and knowledge about friction coefficients are not an approach which would be considered valid in a developmental setting. Instead, it is desirable that the robot learns what stability refers to from its own experience, similar to a child who learns to associate certain actions with consequences on the external world, such as the pressure exerted on the palm while steadily holding an object. Unfortunately, the tactile sensors needed for this kind of measurements were not available on the robotic platform iCub. Thus, we could not make use of information provided e.g. by pressure sensors to infer the grasp stability. Instead, this information was provided from the outside, i.e. the experimenter categorized each performed grasp as stable, unstable, or failed. The objects were selected to represent a wide range of features, but only objects which were in principle – with one grasp or another – suitable to be held by iCub's hand were used for the final data collection. Each action, i.e. probe and grasp combination, was applied to each object between three and ten times. A total of 758 trials were collected.

Examples of the objects used in the experiments are shown in Figure 13.

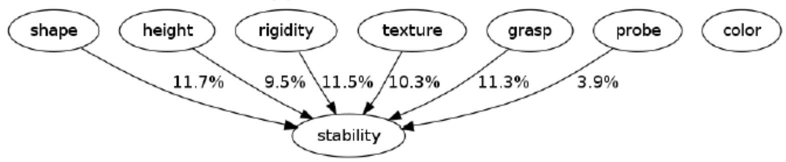
Figure 13: The set of objects employed for the experiments (see text for details).

In summary, in a first stage, data were used to learn the structure of the Bayesian Network using three different methods (for comparison): pure Dynamic Programming (DP) and a hybrid approach using DP to restrict the search space which is subsequently sampled using Markov Chain Monte Carlo methods (DPMCMC). Both of these approaches are subsumed under the notion of “DP methods" in the following. All algorithms here made use of interventional data, assuming perfect interventions. Structure learning using the K2 algorithm was also performed. K2 is a local, greedy search algorithm in the space of directed acyclic graphs. The results compared with manual design of the BN structure is shown below.



(a) Topology uncovered by DP and DPMCMC

(b) Topology uncovered by K2



(c) Manually designed topology

The subsequent step is to learn the network parameters (conditional probabilities). The internal parameters of the nodes were approximated through MAP estimation using *BDeu* priors initialized with a weight of 10 percent of the whole data set, which amounts to 7.5 pretended observations for each unseen event. Finally, as it is often the case with BNs, there is no ground truth on which to evaluate the inference capabilities of any learned network. Yet, it is possible to judge the relative performances of different network topologies using cross validation. In our particular case, we are interested in how good the predicted stability matches the observed stability for any of the computed networks. Note that it would be possible to infer the probability distribution over action features from known object features and stability outcomes as well as any other combination of random variables (RVs). However, as prediction capabilities of the networks form the basis for the later computation of affordances, we are going to limit ourselves to the evaluation of the stability inference problem. To do this, leave one out cross validation was used. The parameters of the three graph topologies shown in the previous figure were trained on the data from all experimental trials except those trials involving the object used for eventual inference evaluation. Then, each network was queried for a probability distribution Q over stability outcomes for each object/action pairing. From the data spared from parameter estimation, the MLE estimate of the distribution over stability outcomes P was computed. After that, the two distributions P and Q were compared using the KL divergence.

In short, statistical tests show ($p<0.01$) that the three methods perform differently. DPBN has the higher predictive power (for grasp stability), the manually designed networks lays somewhere in between and the K2 method is the worst.

In a final experiment – now that the BNs performance has been assessed – we tried to formalize affordances as the “best” match between object and action. The goal is to arrive at a scoring function measuring the match between each action and object, i.e. the affordance of the object formalized as *P(a|o)*. As mentioned earlier this probabilistic formalization is not only important in its own right, but also integrates into models of the mirror neuron system (Metta et al.). It is to be noted that both the distribution over the possible actions can be computed (though expensive) but also (more simply) only the maximum of such distribution.

As mentioned earlier, this model of affordances is well suited to be integrated in a model of mirror neurons (action recognition) for the iCub where affordances act as priors in a Bayes classification problem. This work can further be integrated (although this is left as a future extension) with the method developed by IST for estimating grasp points on objects.

5 Conclusions

This deliverable presents results from experiments in neuroscience, developmental psychology and experimental robotics aiming at the understanding of the complex cognitive processes underlying the acquisition and exploitation of objects affordances. The presented work has laid out a basis for:

- Modelling and acquisition of object's affordances
- Dealing both with learning from self-experience and by observation,
- Practical implementation of learning and execution algorithms in real setups.
- Representing a natural link between lower level sensorimotor coordination levels and higher level imitation and communication levels.
- The model proves to be general enough to incorporate additional actions such as grasping or to be used for learning verbal descriptions (word-label associations).

These results are the ground for the future comparison between the computational model and neural and behavioural data which may provide further insights as to complete and extend the model itself., Effort will be put on the development of sequential method for constantly adapting the affordances network and illustrate the utilization of the network for planning, recognition and communication tasks.

6 References

[1] James J. Gibson. The Ecological Approach to Visual Perception. Boston: Houghton-Mifflin. 1979.

[2] Donald Norman, The Design of Everyday Things. Basic Books. 1988.

[3] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll. Understanding and Sharing Intentions: The origins of cultural cognition. Behavioral and Brain Sciences, 2004.

[4] Marcel Brass and Cecilia Heyes. Imitation: Is cognitive neuroscience solving the correspondence problem?

[5] Visual Learning by Imitation With Motor Representations, Manuel Lopes and José Santos-Victor, IEEE Transactions on System Man and Cybernetics - Part B: Cybernetics, June 2005.

[6] From sensorimotor development to object perception. L. Natale, F. Orabona, F. Berton, G. Metta, G. Sandini. In the International Conference of Humanoids Robotics. 2005.

[7] L. Natale, F. Orabona, G. Metta, G. Sandini. Sensorimotor coordination in a humanoid robot: learning about objects through grasping. Accepted for publication into International Journal of Humanoid Robotics. 2006.

[8] Clark Glymour, Learning, prediction and causal Bayes nets, Trends in Cognitive Sciences, Vol.7 No.1, January 2003.

[9] M. Schenatti, L. Natale, G. Metta, and G. Sandini, Object Grasping Data-Set. Genova, Italy: Lira-Lab., Univ. Genova, 2003.

[10] Kevin P. Murphy, Inference and Learning in Hybrid Bayesian Networks, Technical Report UCB/CSD-98-990, EECS Department, University of California, Berkeley, 1998.

[11] G di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G Rizzolatti. Understanding motor events: a neurophysiological study. Experimental Brain Research, 91:176-180, 1992.

[12] L. Fogassi, V. Gallese, L. Fadiga, G. Luppino, M. Matelli and G. Rizzolatti. Coding of peripersonal space in inferior premotor cortex (area F4). J Neurophysiol, 76:141-57, 1996.

[13] V. Gallese, L. Fadiga, L. Fogassi and G. Rizzolatti. Action recognition in the premotor cortex. Brain, 119:593-609, 1996.

[14] M.C. Hepp-Reymond, E.J. Husler, M.A. Maier and H.X. Ql. Force-related neuronal activity in two regions of the primate ventral premotor cortex. Can J Physiol Pharmacol, 1994; 72: 571-9.

[15] E. Kohler, C.M. Keysers, A. Umiltà , L. Fogassi, V. Gallese, and G. Rizzolatti. Hearing sounds, understanding actions: Action representation in mirror neurons. Science, 297:846-848, 2002.

[16] A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, and G. Rizzolatti. Object representation in the ventral premotor cortex (area F5) of the monkey. J Neurophysiol. 1997;78:2226-30.

[17] G. Rizzolatti and L. Fadiga. Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). In: G. R. Bock & J. A. Goode (Eds.), Sensory Guidance of Movement, Novartis Foundation Symposium (pp. 81-103). Chichester: John Wiley and Sons, 1998.

[18] G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli.. Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. Experimental Brain Research, 71:491–507, 1988.

[19] G. Rizzolatti and L. Craighero. The mirror-neuron system. Ann Rev Neurosci, 27:169-92, 2004.

[20] G. Rizzolatti, L. Fadiga, V. Gallese and L. Fogassi.. Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3:131-141, 1996.

[21] M.A Umiltà, E.Kohler, V.Gallese, L. Fogassi, L. Fadiga, C. Keysers and G. Rizzolatti. I know what you are doing. a neurophysiological study. Neuron, 31:155-65, 2001.

[22] C. von Hofsten and L. Rönnqvist. Preparation for grasping an object: A developmental study. Journal of Experimental Psychology: Human Perception and Performance, pages 610-621, 1988.

[23] C. von Hofsten and S. Fazel-Zandy. Development of visually guided hand orientation in reaching. Journal of Experimental Child Psychology, 38:208-219, 1984.

[24] C. von Hofsten and K. Johansson. Planning to reach for a rotating rod: Developmental aspects, volume Manuscript. 2005.

[25] J. J. Lockman, D. H. Ashmead, and E. W. Bushnell. The development of anticipatory hand orientation during infancy. Journal of Experimental Child Psychology, 37:176-186, 1984.

[26] H. Örnkloo and C. von Hofsten, Fitting objects into holes: on the development of spatial cognition skills. Developmental Psychology. 2006

[27] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and José Santos-Victor, “Learning Object Affordances: From Sensory Motor Maps to Imitation, “IEEE Transactions on Robotics, vol. 24(1), February 2008.

[28] L. Montesano and M. Lopes. Learning grasping affordances from local visual descriptors. IEEE International Conf.Development and Learning, Shanghai, China, 2009

[29] V. Krunic, G. Salvi, A. Bernardino, L. Montesano, J. Santos-Victor Affordance Based Word-To-Meaning Association,. International Conference on Robotics and Automation, 2009, Kobe, Japan .

[30]V. Krunic, G. Salvi, A. Bernardino, L. Montesano, J. Santos-Victor. Associating word descriptions to learned manipulation task models., IROS-2008 WORKSHOP on Grasp and Task Learning by Imitation.

[31] Patrice Senot, Laila Craighero, Giovanni Ottoboni, Chiara Zingales, Charalambos Papaxanthis, Luciano Fadiga. *Corticospinal excitability during motor preparation: a comparison among execution, voluntary inhibition, and observation of an interceptive task.* Submitted to European Journal of Neuroscience.

[32] Ellmer, A. Probabilistic Modeling of Grasping Affordances in Robots using Bayesian Networks. Bachelor Thesis in Cognitive Science, Institute for Cognitive Science, University of Osnabruck, Germany.

[33] Saxena, A., Driemeyer, J., and Ng, A. Y. Robotic grasping of novel objects using vision. International Journal of Robotics Research (IJRR), 2008..