

2010 Special Issue

The iCub humanoid robot: An open-systems platform for research in cognitive development

Giorgio Metta^a, Lorenzo Natale^a, Francesco Nori^a, Giulio Sandini^a, David Vernon^a, Luciano Fadiga^{a,b}, Claes von Hofsten^c, Kerstin Rosander^c, Manuel Lopes^d, José Santos-Victor^e, Alexandre Bernardino^e, Luis Montesano^e

^a Italian Institute of Technology (IIT), Italy

^b University of Ferrara, Italy

^c University of Uppsala, Sweden

^d University of Plymouth, UK

^e Instituto Superior Técnico, Portugal

article info

Keywords:

Open-source humanoid robot
Cognition
Development
Learning
Phylogeny
Ontogeny

abstract

We describe a humanoid robot platform the iCub which was designed to support collaborative research in cognitive development through autonomous exploration and social interaction. The motivation for this effort is the conviction that significantly greater impact can be leveraged by adopting an open systems policy for software and hardware development. This creates the need for a robust humanoid robot that offers rich perceptuo-motor capabilities with many degrees of freedom, a cognitive capacity for learning and development, a software architecture that encourages reuse & easy integration, and a support infrastructure that fosters collaboration and sharing of resources. The iCub satisfies all of these needs in the guise of an open-system platform which is freely available and which has attracted a growing community of users and developers. To date, twenty iCubseach comprising approximately 5000 mechanical and electrical parts have been delivered to several research labs in Europe and to one in the USA.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Robotics, by definition, takes inspiration from nature and the humanoid concept is perhaps the best example. When we consider the possibility of creating an artefact that acts in the world, we face a preliminary and fundamental choice: efficiency (achieved by being task-specific) or versatility (achieved by biological-compatibility development). The first option leads to the realization of automatic systems that are very fast and precise in their operations. The limitations of automatic systems are purely technological ones (e.g. miniaturization). The second option is what we consider to be a humanoid: a biological-like system which takes decisions and acts in the environment, which adapts and learns how to behave in new situations, and which invents new solutions on the basis of the past experience. The fascinating aspect of the humanoid is the possibility to interact with it: to teach, to demonstrate, even to communicate. It should be stressed that the attempt to adopt the strategy of 'biological compatibility' does not

represent an intellectual exercise but is prompted by the idea that a humanoid interacting with human beings must share with them representations, motor behaviours and perhaps, even kinematics and degrees of freedom.

To interact, a humanoid must first act (and not simply move), perceive, categorize and therefore, understand. These capabilities cannot arise from pre-compiled software routines. On the contrary, they realize themselves through an ontogenetic pathway, simulating what happens in developing infants. In other words, humanoids must act in the environment to know it. It should be stressed that 'to know the environment' does not mean to categorize an assembly of static structures and objects but requires, as an essential requisite, to understand the consequences of generated actions (e.g. a glass breaks when it falls on the ground). During this

Corresponding author. Tel.: +39 01071781411.

E-mail addresses: pasa@liralab.it, giorgio.metta@iit.it (G. Metta).

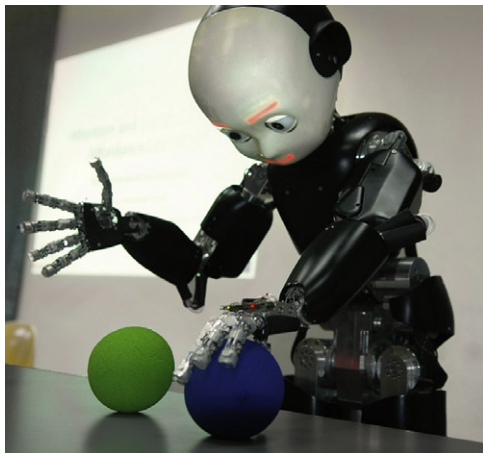


Fig. 1. The iCub humanoid robot: an open-systems platform for research in cognitive development.

a new discipline in which bi-directional benefits are expected. In fact, this knowledge sharing rewards not only robotics but also neuroscience since the developing (learning) humanoid forms a behaving model to test neuro-scientific hypotheses by simplifying some extremely complex problems. Particularly, it allows what is not conceivable in human neuroscience: to investigate the effects of experimental manipulations on developmental processes. This opens up vast new opportunities for advancing our understanding of humans and humanoids.

This paper describes the development of the iCub humanoid robot (see Fig. 1) and our efforts to navigate this uncharted territory, aided by a constantly growing community of iCub users and developers.

The iCub is a 53 degree-of-freedom humanoid robot of the same size as a three or four year-old child. It can crawl on all fours and sit up. Its hands allow dexterous manipulation and its head and eyes are fully articulated. It has visual, vestibular, auditory, and haptic sensory capabilities. The iCub is an open systems platform: researchers can use it and customize it freely since both hardware and software are licensed under the GNU General Public Licence (GPL).¹

The iCub design is based on a road map of human development (von Hofsten, Fadiga, & Vernon, *in press*) (see Section 3). This description of human development stresses the role of prediction into the skillful control of movement: development is in a sense the gradual maturation of predictive capabilities. It adopts a model of "sensorimotor" control and development which considers "action" (that is, movements with a goal, generated by a motivated agent which are predictive in nature) as the basic element of cognitive behaviours. Experiments with infants and adults have shown that the brain is not made of a set of isolated areas dealing with perception or motor control but rather that multisensory neurons are the norm. Experiments have proven the involvement of the motor system, including the articulation of speech, in the fine perception of the movements of others. The iCub employs a computational model of affordances which includes the possibility of learning both the structure of dependences between sets of random variables (e.g. perceptual qualities vs. action and results), their effective links and their use in deciding how to control the robot. Affordances form the quintessential primitives of cognition by mixing perception and action in a single concept or representation. It builds on a computational model of imitation and interaction between humans and

robots by evaluating the automatic construction of models from experience (e.g. trajectories), their correction via feedback, timing and synchronization. This explores the domain between mere sensorimotor associations and the possibility of true communication between robot and people. The iCub involved the design from scratch of a complete humanoid robot including mechanics, electronics (controllers, I/O cards, buses, etc.) and the related firmware and it adopted and enhanced open-systems middleware (YARP) (Metta, Fitzpatrick, & Natale, 2006). Finally, it has resulted in the creation of a community of active users and researchers working on testing, debugging, and improving the iCub of the future.

2. Design goals

The design of the iCub started from the consideration that the construction of cognitive systems could not progress without a certain number of ingredients: the development of a sound formal understanding of cognition (Vernon, Metta, & Sandini, 2007), the study of natural cognition and, particularly important, the study of the development of cognition (Sandini, Metta, & Konczak, 1997; von Hofsten, 2003), the study of action in humans by using neuroscience methods (Fadiga, Fogassi, Gallese, & Rizzolatti, 2000; von Hofsten, 2004), and the physical instantiation of these models in a behaving humanoid robot (Metta, Sandini, & Konczak, 1999; Metta, Sandini, Natale, & Panerai, 2001).

Our research agenda starts from cognitive neuroscience research and proceeds by addressing, for example, the role of manipulation as a source of knowledge and new experience, as a way to communicate socially, as a tool to teach and learn, or as a means to explore and control the environment. We would like to stress here that collaboration between neuroscience, computer science, and robotics is truly intended as bi-directional. On one side, the iCub cognitive architecture is a system as much as possible "biologically oriented".² On the other side, real biological systems were examined according to problems that we deemed important for elucidating the role of certain behaviours or brain regions in a larger picture of the brain. Examples of this research are: the ability to grasp unknown objects on the basis of their shape and position with one and two hands, to assemble simple objects with plugs, and to coordinate the use of two hands (e.g. parts mating, handling of soft materials). These abilities require visuo-haptic object recognition and multimodal property transfer, visual recognition of the body gestures of others, imitation of one and two-hand gestures, and communication and interaction through body and hand gestures.

A no-less-important scientific objective is the study of the initial period of human cognitive development and its implementation on the iCub. Our working method is, in fact, not to pre-program the cognitive skills outlined earlier but, similarly to what happens in humans, to implement them into a system that can learn much like a human baby does. We understand aspects of human development and can make specific informed choices in building an artificial adaptable system. For example, developmental science now points out at how much action, perception and cognition are tightly coupled in development. This means that cognition cannot be studied without considering action and embodiment and how perception and cognition are intertwined into

¹ The iCub software and hardware are licensed under the GNU General Public Licence (GPL) and GNU Free Documentation Licence (FDL), respectively.

² It is important to note that biological plausibility or similarity in the iCub is not intended as a faithful implementation of neural simulations to a very detailed level. We don't think that this approach is feasible given the available hardware. The digital computer is not the brain and it would be wasteful to try to use computers in this sense. On the other hand, the gross features of the architecture are biologically plausible by including attention, memory (procedural and declarative), reaching, grasping, action selection, and affective states.

development (von Hofsten, 2004). Exemplar experimental scenarios are discovering the action possibilities of the body (the so called body map), learning to control one's upper and lower body (crawling, bending the torso) to reach for targets, learning to reach static and moving targets, and learning to balance in order to perform stable object manipulations when crawling or sitting. They include also discovering and representing the shape of objects and discovering and representing object affordances (e.g. the use of "tools"). Interaction with other agents is also important: recognizing manipulation abilities of others and relating those to one's own manipulation abilities, learning to interpret and predict the gestures of others, learning new motor skills and new object affordances by imitating manipulation tasks performed by others, learning what to imitate and when to imitate others gestures, and learning regulating interaction dynamics. Clearly, this is an ambitious research programme and it is far from being completed. However, we have set the basis for a solid development in this direction by providing the platform and by setting up the whole infrastructure (together with examples and large parts of this set of behaviours).

To enable the investigation of relevant cognitive aspects of manipulation the design was aimed at maximizing the number of degrees of freedom (DOF) of the upper part of the body (head, torso, arms, and hands). The lower body (legs) were initially designed to support crawling "on four legs" and sitting on the ground in a stable position (and smoothly transition from crawling to sitting). A recent study and consequent modification of the legs allows bipedal walking, although this is still theoretical since the control software has not been developed yet. We are also designing a mobile base (on wheels) for the iCub which will allow mobility and autonomy (on battery). Mobility, in general, whether on wheels or by crawling, allows the robot to explore the environment and to grasp and manipulate objects on the floor. The size of the iCub is that of a three- to four-year-old child and the total number of degrees of freedom for the upper body is 41 (7 for each arm, 9 for each hand, 6 for the head and 3 for the torso and spine). The sensory system includes binocular vision, touch, binaural audition and inertial sensors. Functionally, the iCub can coordinate the movement of the eyes & hands, grasp and manipulate lightweight objects of reasonable size and appearance, crawl on four legs and sit (Metta, Sandini, Vernon, Natale, & Nori, 2008; Tsagarakis et al., 2000). Such a tool did not exist prior to the construction of the iCub even considering the humanoid robotic products developed recently by Japanese companies (e.g. Sony, Honda, etc.) and it is still the only complete open-systems humanoid robot available today.

To emphasize again, the design of the iCub places strong emphasis on manipulation since neural science tells us a story – a summary can be found in Arbib (2000) in which manipulation is central to human cognition. In fact, manipulation is the way through which we get to grips with the world, with the concept of objecthood, with the social environment, and further, if we subscribe to this story, communication to the level of language evolved out of a process of adaptation of the manual system into the one that controls speech. Equally important, the iCub has legs for crawling which give the robot a chance for building its own experience by exploring the environment, fetching objects, etc. This raises a whole new set of issues since the robot has to link the frame of reference of its perceptual abilities to a moving environment rather than to the usual fixed one as in many stationary platforms. One example is in building the understanding of the limits of the robots own body: in this case, the robot can exploit the fact that its body is relatively constant over time while the environment has a higher variability. A high variability in the environment helps in building this important distinction.

3. Foundations of human development

Our goal in studying the development of early cognition in humans is to model the relevant aspects of such a process within the

boundaries of an artificial system. In particular, we investigate the time frame of a developmental process that begins to guide action by internal representations of upcoming events, by the knowledge of the rules and regularities of the world, and by the ability to separate means and end (or cause and effect). We study and model how young children learn procedures to accomplish goals, how they learn new concepts, and how they learn to improve plans of actions. This research is strongly driven by studies of developmental psychology and cognitive neuroscience and it has resulted in a physical implementation on the iCub as well as a road map for the development of cognitive abilities in humanoid robots (von Hofsten et al., in press). To a large extent, this road map is a conceptual framework that forms the foundation of the iCub project. It surveys what is known about cognition in natural systems, particularly from the developmental standpoint, with the goal of identifying the most appropriate system phylogeny and ontogeny. It explored neuro-physiological and psychological models of some of these capabilities, noting where appropriate architectural considerations such as sub-system interdependencies that might shed light on the overall system organization. It uses the phylogeny and ontogeny of natural systems to define the innate skills with which the iCub must be equipped so that it is capable of ontogenetic development, to define the ontogenetic process itself, and to show exactly how the iCub should be trained or to what environments it should be exposed in order to accomplish this ontogenetic development. Finally, it embraces the creation and implementation of an architecture for cognition: a computational framework for the operational integration of the distinct capabilities and cognitive skills developed in the project (these will be discussed in the following sections).

The iCub project takes an enactive approach to the study of cognition whereby a cognitive system develops its own understanding of the world around it through its interactions with the environment (Maturana , 1970, 1975; Maturana & Varela, 1980, 1987; Paolo, Rohde, & Jaeger, 2008; Thompson, 2007; Varela, 1979, 1992) and for which ontogenetic development is the only possible solution to the acquisition of epistemic knowledge (the systems representations). In the enactive approach, cognition is self-organizing and dynamical and corresponds to the acquisition (and development) of anticipatory abilities and the development of an increasing space of interaction between the cognitive agent and its environment. We take this approach also in interpreting cognition in biological systems. Consequently, the next important question is about the principles that govern the ontogenetic development of biological organisms (e.g. as in von Hofsten (2004)). Converging evidence from various disciplines including developmental psychology and neuroscience is showing that behaviour in biological organisms is organized in primitives that we can call actions (as distinct from movements or reactions). Actions are behaviours initiated by a motivated subject, defined by goals and guided using prospective information (prediction). Elementary behaviours are thus not reflexes but actions with goals, where perception and movement are integrated, and they are initiated because of a motivation and that are guided by and through prediction (von Hofsten, 2004).

To make this more operational and provide a description of human development, we have to consider three basic elements:

1. What is innate, where do we start from?
2. What drives development?
3. How is new knowledge incorporated, i.e. what are the forces that drive development?

In looking at the first question, developmental psychologists, typically refer to innate elements in terms of prenatal prestructuring or the so-called core abilities. Neither is to be imagined like a rigid determination of perception action couplings but rather a

means to facilitate development. Examples can be found in the pre-structuring of the morphology of the body, in the perceptual, and in the motor systems.

The motor system requires constraints in order to reduce the large number of effective degrees of freedom and these constraints come in the form of muscular synergies. That is, to facilitate control, the activation of muscles is therefore organized into functional synergies at the beginning of life (and they are probably formed already prenatally ([de Vries, Visser, & Prechtl, 1982](#))). Similarly, perceptual structuring begins early in ontogenesis by relying on the interaction between genetic and self-activity factors ([Johnson, 1997](#); [Quartz & Sejnowski, 1997](#); [von der Malsburg & Singer, 1988](#)). In addition to these, prestructuring comes also in the form of specific core abilities. [Spelke \(2000\)](#) is one of the proponents of this view. She discusses various aspects that show prestructuring, such as the perception of objects and the way they move, the perception of geometric relationships and numerosities, and the understanding of persons and their actions. An important part of the core knowledge has to do with people.

Knowing the initial state of the system is only the first step. A model of human development then requires establishing what causes it. Motivations come in different forms in the newborn: social and explorative. The social motive is what puts the infant in the broader context of other human beings, thus providing further possibilities for learning, safety, comfort, etc. Communication and language also develop within the context of social interaction ([Johnson & Morton, 1991](#)).

The third basic element of this summary of human development is to show how new knowledge is acquired and incorporated. The brain is only one side of this process: without interaction with the environment it would be of little use. Undoubtedly, the brain has its own dynamics (proliferation of neurons, maps formation, migration, etc.) but the final product is shaped by the dynamical interaction with the environment. Factors like exposure or deprivation to the environment, the body biomechanics and body growth are all fundamental to the development of cognition. For instance, the appearance of reaching depends critically on the appearance of 3D perception through binocular disparity, on the emergence of postural control (and muscle strength), on the separation of the extension flexion synergies in the arm and hand, on the perception of external motion, control of the eyes for tracking and so forth. This is to say that no single factor determines the appearance of a particular new behaviour and it is therefore important to model complete systems in order to analyze even relatively simple cognitive behaviours.

Complementary to developmental studies, neurophysiology is also helping to show the inextricably complexity of the brain. Tantalizing results from neuroscience are shedding light on the mixed motor and sensory representations used by the brain during reaching, grasping, and object manipulation. We now know a great deal about what happens in the brain during these activities, but not necessarily why. Is the integration we see functionally important, or just a reflection of evolution's lack of enthusiasm for sharp modularity? A useful concept to help understand how such capabilities could develop is the well-known theory of [Ungerleider and Mishkin \(1982\)](#) who first formulated the hypothesis that the brain's visual pathways split into two main streams: the dorsal and the ventral. The dorsal is the so-called "where" pathway, concerned with the analysis of the spatial aspects of motor control. The ventral is related with the "what", that is, the identity of the targets of action. [Milner and Goodale \(1995\)](#) refined the theory by proposing that objects are represented differently during action than they are for a purely perceptual task. The dorsal deals with the information required for action, whereas the ventral is important for more cognitive tasks such as maintaining an object's identity and constancy. Although the dorsal/ventral segregation is

emphasized by many commentators, it is significant that there is a great deal of cross talk between the streams ([Jeannerod, 1997](#)).

Among the arguments in favour of the 'pragmatic' role of the visual information processed in the dorsal stream are the functional properties of the parieto-frontal circuits. For reason of space we cannot review here the functional properties of these circuits, e.g. that formed by area LIP and FEF, those constituted of parietal area VIP (ventral intraparietal) and frontal area F4 (ventral premotor cortex) or the pathway that connects area AIP (anterior intraparietal) with area F5 (dorsal premotor cortex). The same functional principle is valid, however, throughout these connections. Area F5, one of the main targets of the projection from AIP (to which it sends back recurrent connections), was thoroughly investigated by [Rizzolatti and colleagues \(Gallese, Fadiga, Fogassi, & Rizzolatti, 1996\)](#). F5 neurons can be classified in at least two different categories: canonical and mirror.

Canonical and mirror neurons are indistinguishable from each other on the basis of their motor responses. Their visual responses, however, are quite different. The canonical type is active in two situations: (1) when grasping an object and (2) when fixating that same object. For example, a neuron active when grasping a ring also fires when the monkey simply looks at the ring. This could be thought of as a neural analogue of the "affordance" of [Gibson \(1977\)](#). The second type of neuron identified in F5, the mirror neuron ([Fadiga et al., 2000](#)), becomes active under either of two conditions: (1) when manipulating an object (e.g. grasping it, as for canonical neurons), and (2) when watching someone else performing the same action on the same object. This is a more subtle representation of objects, which allows and supports, at least in theory, mimicry behaviours. In humans, area F5 is thought to correspond to Broca's area and there is an intriguing link between gesture understanding, language, imitation, and mirror neurons ([Fadiga, Craighero, Buccino, & Rizzolatti, 2002](#); [Rizzolatti & Arbib, 1998](#)). The STS region and parts of TE contain neurons that are similar in response to mirror neurons ([Perrett, Mistlin, Harries, & Chitty, 1990](#)). They respond to the sight of the hand; the main difference compared to F5 is that they lack the motor response. It is likely that they participate in the processing of the visual information and then communicate with F5 ([Gallese et al., 1996](#)), most likely via the parietal cortex.

Studying the motor system is consequently a complete activity involving sensorimotor loops which have a role in the recognition of objects ([Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997](#)), of actions ([Rizzolatti, Fadiga, Gallese, & Fogassi, 1996](#)), in planning and understanding the intentions of others ([Fogassi et al., 2005](#)) as well as in language ([D'Ausilio et al., 2009](#); [Fadiga et al., 2002](#)). The involvement of the motor areas during observation of actions has been recently analyzed in human subjects using the H-reflex and TMS-evoked motor potentials ([Borron, Montagna, Cerri, & Baldissera, 2005](#); [D'Ausilio et al., 2009](#)). It has been shown that the so-called "motor resonance" phenomenon ([Gallese et al., 1996](#)) is not relegated to the cortex but, rather, it spreads far deeper than initially thought. It has been shown that the spinal cord excitability is modulated selectively under threshold by the observation of others. In particular, in this experiment, the excitability of the spinal cord was assessed and it was determined to reflect an anticipatory pattern similar to the actual muscular activation with respect to the kinematics of the action.

These studies in neuroscience provided the requirements and boundary condition for the design and implementation of the iCub cognitive architecture. This architecture was initially loosely modelled after the "global workspace architecture" of [Shanahan \(2005a, 2005b, 2006\)](#) but later evolved into something different which is unique to the iCub.

This work on neuroscience was complemented by other studies in developmental psychology which culminated in a road map for

the development of cognitive abilities in humanoid robots based on the ontogeny of human neonates. This road map also defines a set of scenarios and empirical tests for the iCub cognitive architecture. The main idea is to be able to test the iCub in the same manner as a developmental psychologist would test an infant in a laboratory experiment.

4. Specific results

In this section, we summarize the main results to convey some of the most exciting features of the iCub. We begin by describing briefly the physical iCub platform and its software architecture before focussing on sensorimotor coordination, manipulation and affordances, and imitation & communication.

4.1. Mechatronics of the iCub

The iCub is approximately 1m tall and weighs 22 kg. From the kinematic and dynamic analysis, the total number of degrees of freedom for the upper body was set to 38 (7 for each arm, 9 for each hand, and 6 for the head). The hands each have three independent fingers and the fourth and fifth to be used for additional stability and support (only one DOF overall). They are tendon driven, with most of the motors located in the forearm. For the legs the simulations indicated that for crawling, sitting and squatting a 5 DOF leg is adequate. However, it was decided to incorporate an additional DOF at the ankle to support standing and walking. Therefore each leg has 6 DOF: these include 3 DOF at the hip, 1 DOF at the knee and 2 DOF at the ankle (flexion/extension and abduction/adduction). The foot twist rotation was not implemented. Crawling simulation analysis also showed that for effective crawling a 2 DOF waist/torso is adequate. However, to support manipulation a 3 DOF waist was incorporated. A 3 DOF waist provides increased range and flexibility of motion for the upper body resulting in a larger workspace for manipulation (e.g. when sitting). The neck has a total of 3 DOF and provides full head movement. The eyes have further 3 DOF to support both tracking and vergence behaviors.

From the sensory point of view, the iCub is equipped with digital cameras, gyroscopes and accelerometers, microphones, and force/torque sensors. A distributed sensorized skin is under development using capacitive sensors technology. Each joint is instrumented with positional sensors, in most cases using absolute position encoders. A set of DSP-based control cards, custom-designed to fit the iCub, takes care of the low-level control loop in real-time. The DSPs communicate with each other via a CAN bus. Four CAN bus lines connect the various segments of the robot. All sensory and motor-state information is transferred to an embedded Pentium based PC104 card that handles synchronization and reformatting of the various data streams. Time consuming computation is typically carried out externally on a cluster of machines. The communication with the robot occurs via a Gbit Ethernet connection.

The iCub is equipped with an umbilical cord which contains both an Ethernet cable and power to the robot. At this stage there is no plan for making the iCub fully autonomous in terms of power supply and computation (e.g. by including batteries and/or additional processing power on board).

Certain features of the iCub are unique. Tendon driven joints are the norm both for the hand and the shoulder, but also in the waist and ankle. This reduces the size of the robot but introduces elasticity that has to be considered in designing control strategies where high forces might be generated. The hand, for example, is fully tendon-driven (see Fig. 2). Seven motors are placed remotely in the forearm and all tendons are routed through the wrist mechanism (a 2 DOF differential joint). The thumb, index, and middle finger are driven by a looped tendon in the proximal joint.

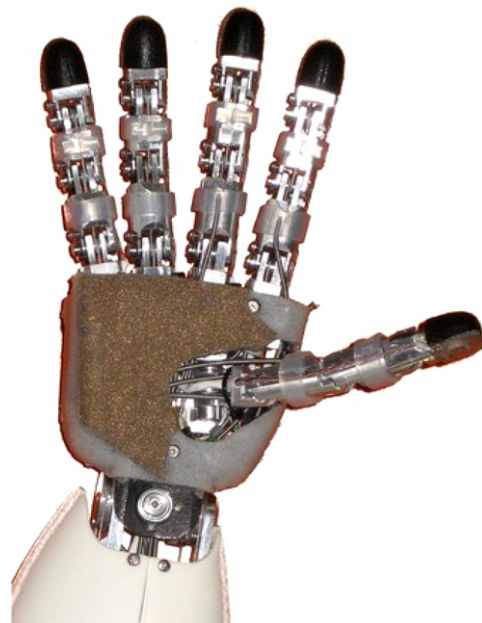


Fig. 2. The hand of the iCub, showing some of the tendons, the sensorized fingertips and the coating of the sensors of the palm (108 taxels overall). Tendons are made of Teflon-coated cables sliding inside Teflon coated flexible steel tubes.

Motion of the fingers is driven by tendons routed via idle pulleys on the shafts of the connecting joints. The flexing of the fingers is directly controlled by the tendons while the extension is based on a spring return mechanism. This arrangement saves one cable per finger. The last two fingers are coupled together and pulled by a single motor which flexes 6 joints simultaneously. Two more motors, mounted directly inside the hand, are used for abduction/adduction movements of the thumb and all fingers except the middle one which is fixed with respect to the palm. In summary, eight DOF out of a total of nine are allocated to the first three fingers, allowing considerable dexterity. The last two fingers provide additional support to grasping. Joint angles are sensed using a custom-designed Hall-effect-magnet pair. In addition room for the electronics and tactile sensors has been planned. The tactile sensors are under development (Maggiali et al., 2008). The overall size of the palm has been restricted to 50 mm in length; it is 34 mm wide at the wrist and 60 mm at the fingers. The hand is only 25 mm thick.

4.2. Software architecture

Considerable effort went into the development of a suitable software infrastructure. The iCub software was developed on top of YARP (Fitzpatrick, Metta, & Natale, 2008). The iCub project supported a major overhaul of the YARP libraries to adapt to a more demanding collaborative environment. Better engineered software and interface definitions are now available. YARP is a set of libraries that supports modularity by abstracting two common difficulties in robotics: namely, modularity in algorithms and in interfacing with the hardware. Robotics is perhaps one of the most demanding application environments for software recycling where hardware changes often, different specialized OSs are typically encountered in a context with a strong demand for efficiency. The YARP libraries assume that an appropriate real-time layer is in charge of the low-level control of the robot and instead takes care of defining a soft real-time communication layer and hardware interface that is suited for cluster computation. YARP also takes care of providing independence from the operating system and the development environment. The main tools in this respect are ACE (Huston, Johnson, & Syiid, 0000) and CMake. The former is

an OS-independent communication library that hides the quirks of interprocess communication across different OSs. CMake is a cross-platform make-like description language and tool to generate appropriate platform specific project files.

YARP abstractions are defined in terms of protocols. The main YARP protocol addresses inter-process communication issues. The abstraction is implemented by the `Port C++` class. Ports follow the observer pattern by decoupling producers and consumers. They can deliver messages of any size, across a network using a number of underlying protocols (including shared memory when possible). In doing so, Ports decouple as much as possible (as function of a certain number of user-defined parameters) the behavior of the two sides of the communication channels. Ports can be commanded at run time to connect and disconnect.

The second abstraction of YARP concerns hardware devices. The YARP approach is to define interfaces for classes of devices to wrap native code APIs (often provided by the hardware manufactures). Change in hardware will likely require only a change in the API calls (and linking against the appropriate library). This easily encapsulates hardware dependencies but leaves dependencies in the source code. The latter can be removed by providing a "factory" for creating objects at run time (on demand). The combination of the port and device abstractions leads to remotable device drivers which can be accessed across a network: e.g. a grabber can send images to a multitude of listeners for parallel processing.

Overall, YARP's philosophy is to be lightweight and to be "gentle" with existing approaches and libraries. This naturally excludes hard real-time issues that have to be necessarily addressed elsewhere, likely at the OS level.

4.3. Sensorimotor coordination models

The iCub's cognitive capabilities depend greatly on the development of sensorimotor coordination and sensorimotor mapping. At the outset, we identified how the sensorimotor system is determined by biology, how this is expressed in development, and how experience enters into the process in forming reliable and sophisticated tools for exploring and manipulating the outside world. Our particular concern here is to identify the sensory information (visual, proprioceptive, auditory) that is necessary to organize goal-directed actions. As with everything else, these issues are first investigated in humans and then used to define the iCub's cognitive architecture. The research on sensorimotor coordination has two distinct themes.

1. Modelling how sensorimotor systems evolve from sets of relatively independent mechanisms to unified functional systems. In particular, we study and model the ontogenesis of looking and reaching, for example by asking the following questions: how does gaze control evolve from the saccadic behaviour of newborns to the precise and dynamic mode of control that takes into account both the movement of the actor and the motion of objects in the surrounding? How does reaching evolve from the crude coordination in newborns to the sophisticated and skillful manipulation in older children? In addition, we model how different sensorimotor maps (for gaze/head orienting, for reaching, for grasping, etc.) can be fused to form a subjectively unitary perception/action system. We look also at how the brain coordinates different effectors to form a "pragmatic" representation of the external world using neurophysiological, psychophysical, and robotics techniques.
2. Modelling the role of motor representation as tools serving not only action but also perception. This topic, on which we will expand later in the paper, clearly benefits from a unifying vision based on the idea that the motor system (at least at its representational level) forms the "active filter" carving out passively perceived stimuli by means of attentional or "active perception" processes.

The postulate that action and perception are interwoven with each other and form the basis of higher cognition is in contrast with the established modular view according to which perceptually-related activity in motor systems could still be accounted for in the sense of bottom up effects. As the importance of sensory input on the control of actions is widely agreed upon, an evaluation of, and, eventually, decision between, the two alternative positions critically depends on the question whether activity in motor systems is relevant for perception and comprehension.

In summary, along these lines we realized a layered controller system for the iCub including:

1. Spinal behaviours: e.g. rhythmic movement and basic synergies, force feedback. We developed an architecture for the generation of discrete and rhythmic movements where trajectories can be modulated by high-level commands and sensory feedback (Degallier et al., 2008).
2. Eye movements and attention: an attention system was developed which includes sensory input processing (vision and audition), eye neck coordination, eye movements (smooth pursuit, saccades, VOR and vergence). Methods for tracking behind occlusions have been also investigated (Ruesch, Lopes, Hornstein, Santos-Victor, & Pfeifer, 2008).
3. Reaching and body schema: a robust task-space reaching controller has been developed and methods for learning internal models tested. Specifically, generic inverse kinematics models and human-like trajectory generation has been implemented for the iCub by taking into account various constraints such as joint limits, obstacles, redundancy and singularities (Pattacini, Nori, Natale, Metta, & Sandini, in press).
4. Grasping: finally, based on reaching and orienting behaviours, a grasping module has been implemented. This allows the coordination of looking (for a potential target), reaching for it (placing the hand close to the target) and attempting a grasping motion (or another basic action).

The investigation from a neuroscientific perspective of sensorimotor representations and their role in cognitive functions contributed directly to the implementation of sensorimotor skills in the iCub based on a biologically plausible model for object interaction and the recognition of actions in others. Many experimental techniques and approaches have been used to pursue this goal. In particular, we conducted electrophysiological experiments on both humans and animals (transcranial magnetic stimulation, single neuron recordings), brain imaging experiments (functional magnetic resonance, near infrared spectroscopy), kinematics and gaze tracking recordings, behavioural experiments on both normal individuals and patients (autistic children and frontal aphasic patients). These contributions served to clarify the strict interdependence between the motor command and the sensory consequences of action execution and its fundamental role in the building and development of cognitive functions.

For example, functional brain studies showed that the human mirror system responds similarly to the primate mirror neuron system, and relies on an inferior frontal, premotor, and inferior parietal cortical network. Furthermore, this mirror system is more activated when subjects observe movements for which they have developed a specific competence or when they listen to rehearsed musical pieces compared with music they had never played before. Though humans rely greatly on vision, individuals who lack sight since birth still retain the ability to learn actions and behaviours from others. To what extent is this ability dependent on visual experience? Is the human mirror system capable of interpreting nonvisual information to acquire knowledge about others? It turns out that the mirror system is also recruited when individuals receive sufficient clues to understand the meaning of the occurring action with no access to visual features, such as when they only

listen to the sound of actions or to action-related sentences. In addition, neural activity in the mirror system while listening to action sounds is sufficient to discriminate which of two actions another individual has performed. Thus, while these findings suggest that mirror system may be activated also by hearing, they do not rule out that its recruitment may be the consequence of a sound-elicited mental representation of actions through visually-based motor imagery.

We used functional magnetic resonance imaging (fMRI) to address the role of visual experience on the functional development of the human mirror system. Specifically, we determined whether an efficient mirror system also develops in individuals who have never had any visual experience. We hypothesized that mirror areas that further process visually perceived information of others' actions and intentions are capable of processing the same information acquired through nonvisual sensory modalities, such as hearing. Additionally, we hypothesized that individuals would show a stronger response to those action sounds that are part of their motor repertoire.

To this purpose, we used an fMRI sparse sampling six-run block design to examine neural activity in blind and sighted healthy volunteers while they alternated between auditory presentation of hand-executed actions (e.g., cutting paper with scissors) or environmental sounds (e.g., rainstorm), and execution of a "virtual" tool or object manipulation task (motor pantomime). Results show that in congenitally blind individuals, aural presentation of familiar actions compared with the environmental sounds elicited patterns of neural activation involving premotor, temporal, and parietal cortex, mostly in the left hemisphere, similar to those observed in sighted subjects during both aural and visual presentation. These findings demonstrate that a left premotor-temporo-parietal network subserves action perception through hearing in blind individuals who have never had any visual experience, and that this network overlaps with the left-lateralized mirror system network that was activated by visual and auditory stimuli in the sighted group. Thus, the mirror system can develop in the absence of sight and can process information about actions that is not visual. Further, the results in congenitally blind individuals unequivocally demonstrate that the sound of an action engages human mirror system brain areas for action schemas that have not been learned through the visual modality.

Along the same line of investigation, we asked whether other people's actions are understood by projecting them onto one's own action programs and whether this mode of control functions in infants. The gaze and hand movements of both adults and infants were measured in two live situations. The task was either to move an object between two places in the visual field or to observe the corresponding action performed by another person. When the subjects performed the action, infants and adults behaved strikingly similar. They initiated the hand and gaze movements simultaneously and gaze arrived at the goal ahead of the hand. When observing such actions, the initiation of the gaze shifts was delayed relative to the observed movement in both infants and adults but gaze still arrived at the goal ahead of the hand. The infants' gaze shifts, however, were more delayed at the start, less proactive at the goal, and showed kinematic variability indicating that this mode of functioning is somewhat unstable in 10-month-old infants. In summary, the results showed that both adults and infants perceive the goal of the action and move gaze there ahead of time, but they did not support the idea of a strict matching of the kinematics between the eye movements carried out when performing and observing actions.

4.4. Object affordances

The term affordance was originally used by Gibson 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.

Building on the sensorimotor coordination, the iCub can also develop the ability to learn the affordances of objects. Specific models of how the primates brain represents affordances were considered (for example the parietal-frontal circuit) as well as results from psychological sciences. Specifically, we investigated what exploratory behaviours support the acquisition of affordances and what is the relevant information (visual, haptic, motor, etc.). We developed a model of the acquisition of object affordances and how the motor information enters into the description of perceptual quantities. In analogy to what is observed in the brain, we also investigated how the definition of purpose (or goal) participates in the representation of the actions an object affords.

Humans learn to exploit object affordances throughout their entire life 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 this, we distinguish two means of acquisition of object affordances: that is, self-exploration (autonomous learning) and by observation (learning from examples). From 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. Therefore, for learning affordances, it is essential to be able to locate objects in the environment and execute goal-directed motor actions over objects. Much of the work on sensorimotor coordination focuses on the development of capabilities for controlling one's 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 movements with respect to sensory information, it can start interacting with objects and understanding its interface how to grab the 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 affordances and interpreting activities. These capabilities have an important relationship with the development of imitation and gesture communication (to be described below).

For learning affordances, we use Bayesian Networks (BN) to model the dependencies between robot actions, object characteristics, and the resulting effects (Montesano, Lopes, Bernardino, & Santos-Victor, 2000). Briefly, a BN 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 . We assumed that the iCub has developed certain skills prior to be able to learn affordance (as described in Section 4.3): a motor repertoire (A), perhaps derived from experience, an object feature repertoire (F) also potentially acquired via object manipulation and the effects (E) resulting from manipulating the environment.

The interaction of the iCub with the environment is therefore formalized in using one action a from A on certain objects with features F (or a subset of them) to obtain effects e from E . This information can be used to estimate the BN structure and parameters using different learning algorithms. These parameters can be updated online as the robot performs more experiments. Also, they

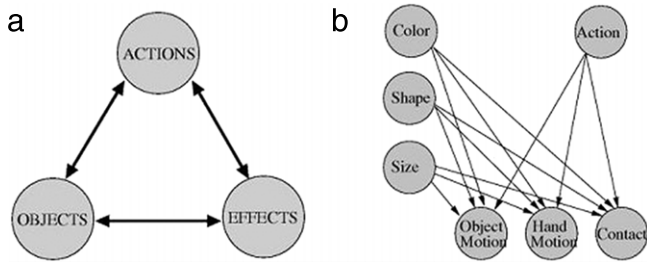


Fig. 3. (a) General affordance scheme relating actions, objects (through their characteristics) and the resulting effects. (b) A particular BN encoding affordances.

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

Inputs	Outputs	Function
(O, A)	E	Predict effect
(O, E)	A	Recognize action & planning
(A, E)	O	Object recognition & selections

can be updated by observation of other agents. Examples are shown in Fig. 3.

This model has some nice properties; for example, affordances can be learned autonomously by experience and by self-observation, restricting the update of the probability distributions. Features can be either selected or ignored, depending on their salience, and the model can be used to perform prediction, recognition, and planning, depending on how the affordance network is traversed. This traversal is 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. Table 1 summarizes some of the basic operations that can be performed with the network.

Based on this previous model, we have performed several experiments with the robotic platform shown in Fig. 4. We used a playground scenario consisting of several objects with two shapes (box and ball), different sizes and colours. The iCub was able to perform three different actions: grasp, tap and touch. An example of an affordance network is shown in Fig. 5. These 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.

4.5. Imitation and communication

Progress has also been made in integrating imitation and communication in an ontogenetic framework on the iCub platform. Imitation plays a central role and communication is strongly related to imitation as regards social cues, turn-taking, and communicative functions. Our particular concern here are the cognitive skills required for imitative behaviours and the cognitive skills required for communicating through body gestures. We also investigated the regulation of interaction dynamics of social interaction during human robot play and its development in ontogeny. The prerequisites for interactive and communicative behaviour grounded

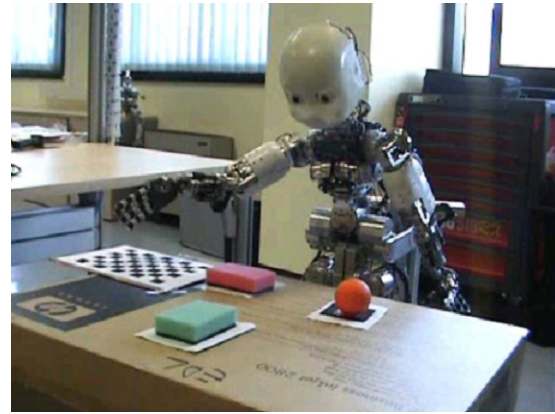


Fig. 4. The playground for the robot contains objects of several sizes, colours and shapes. Protocol: the object to interact with is selected manually, the action is random (from the set of actions A). Object properties are recorded when the hand is not occluding the object. The effects are recorded later and then the robot hand goes open loop to a resting position.

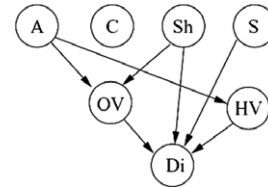


Fig. 5. Learned network. The variables represent A action, C object colour, Sh object shape, S object size, OV object velocity profile, HV hand velocity profile, Di hand object distance profile.

in sensorimotor experience and interaction histories were investigated and developed with specific consideration of interaction kinesics (including gestures, synchronization and rhythms of movements etc.). Social drives for interaction, imitation and communication were considered to make use of non-verbal social cues in ontogeny in the course of human robot interaction.

This work relies on fairly sophisticated cognitive skills which include the ability to recognize and interpret somebody else's gestures in terms of its own capabilities (mirror effects), the ability to learn new gestures on the basis of the observation of those in other individuals, and the ability to recognize the purpose of other people's gestures, such as the goal of manipulating objects in a certain specific way. It also relies on the ability to predict the result of a demonstrated manipulation task and to use this ability to discriminate between good and poor demonstrations of manipulation tasks based on their affordances. Finally, the ability to decide what part of the demonstration is relevant to imitation is required.

Prerequisites to these skills are the skillful control of arms and body in order to produce communicative gestures reflecting communicative timing or turn-taking, tracking and recognizing someone else's gestural timing, synchrony, and social engagement, to generalize and acquire simple communicative behaviours making use of social cues, to respond adequately to timing and gesturing of an interaction partner, and to harness turn taking as the underlying rhythm of gestured communication. That is, both the static aspect of recognition of actions and their social and temporal qualities have to be mastered before proper imitation and communication can happen.

A large part of this iCub work took a human robot interaction perspective to analyzing and developing controllers to enhance human robot communication. This work addressed the above delineated goals of determining the role that timing, social cues, and gesture recognition play in human robot communication. Further,

progress on the development of algorithms for imitation learning was made by extending work on statistical estimate of motion dynamics to allow robust estimation of arbitrary non-linear autonomous dynamical systems.

A number of human studies on various topics pertaining to the basis of human human communication and imitation were also conducted. These studies focused on the observation action/perception action loop for both basic motor tasks and high-level cognitive tasks, such as speech production and perception. In addition, the project conducted a user-study to delineate the variables controlled during imitation of simple goal-directed arm reaching motion. This study informed the development of a computational model of reaching movement that uses the same non-linear dynamical form as that used in the robotics imitation work mentioned above. Further experiments were directed at determining the role of Broca's area in the perception of various types of events (biological vs. non-biological) but also on the involvement of the motor system in the perception of speech and in interpersonal interaction under the influence of a reward.

In particular, one quite fundamental experiment (D'Ausilio et al., 2009) has shown that listening to speech recruits a network of fronto-temporoparietal cortical areas. Classical models consider anterior (motor) sites to be involved in speech production whereas posterior sites are considered to be involved in comprehension. This functional segregation is challenged by action perception theories suggesting that brain circuits for speech articulation and speech perception are functionally dependent. Although recent data show that speech listening elicits motor activities analogous to production, it's still debated whether motor circuits play a causal contribution to the perception of speech.

Here, we set out to investigate the functional contributions of the motor-articulatory systems to specific speech-perception processes. To this end, a cross-over design orthogonalizing the effect of brain-phonology concordance with those of linguistic stimuli and TMS loci was chosen. Phonemes produced with different articulators (lip-related: [b] and [p]; tongue-related: [d] and [t]) were presented in a phoneme discrimination task. The effect of TMS to lip and tongue representations in precentral cortex, as previously described by fMRI, was investigated. Double TMS pulses were applied just prior to stimuli presentation to selectively prime the cortical activity specifically in the lip (LipsM1) or tongue (TongueM1) area. Behavioural effects were measured via reaction times and error rates.

Reaction time performance showed a behavioural double dissociation between stimulation site and stimulus categories. Reaction time change of phonological decisions induced by TMS pulses to either the TongueM1 or LipM1 showed opposite effects for tongue- and lip-produced sounds. Therefore, the stimulation of a given M1 representation led to better performance in recognizing speech sounds produced with the concordant effector compared with discordant sounds produced with a different effector. These results provide strong support for a specific functional role of motor cortex in the perception of speech sounds. In parallel, we tested whether TMS was able to modulate the direction of errors. Errors were grouped in two classes: lip-phoneme errors (L-Ph-miss) and tongue-phoneme errors (T-Ph-miss).

The double dissociation we found in the present work provides evidence that motor cortex contributes specifically to speech perception. As shown by both RTs and errors, the perception of a given speech sound was facilitated by magnetically stimulating the motor representation controlling the articulator producing that sound, just before the auditory presentation. Biologically grounded models of speech and language have previously postulated a functional link between motor and perceptual representations of speech sounds. We demonstrate here for the first time a specific causal link for features of speech sounds. The relevant areas in motor cortex seem to be also relevant for controlling the tongue and lips, respectively.

5. Conclusion

To the best of our knowledge, the iCub cognitive humanoid robot is at the forefront of research in developmental robotics. The iCub was designed completely from scratch mechanics, electronics, firmware, and software specifically with the requirements of developmental cognition in mind. Its design is based on a road map of human development (von Hofsten et al., in press) which already contains a full-fledged program of empirical research that may keep scientists busy for many years to come. This description of human development stresses the role of prediction into the skillful control of movement: development is in a sense the gradual maturation of predictive capabilities. It incorporates a model of sensorimotor control and development which considers action (that is, movements with a goal, generated by a motivated agent which are predictive in nature) as the basic element of cognitive behaviours. Experiments with infants and adults have shown that the brain is not made of a set of isolated areas dealing with perception or motor control but rather that multisensory neurons are the norm. Experiments have proven the involvement of the motor system in the fine perception of others movements including speech. The iCub uses a computational model of affordances which includes the possibility of learning both the structure of dependences between sets of random variables (e.g. perceptual qualities vs. action and results), their effective links and their use in deciding how to control the robot. Affordances are the quintessential primitives of cognition by mixing perception and action in a single concept (representation); this representation has facilitated the creation of a computation model of imitation and interaction between humans and robots by evaluating the automatic construction of models from experience (e.g. trajectories), their correction via feedback, timing and synchronization. This explores the domain between mere sensorimotor associations and the possibility of true communication between robot and people. Finally, the iCub project has given rise to a large and growing community of highly-active users, developers, and researchers drawn from many disciplines, all committed to creating the iCub of the future.

Although much is still to be done to implement the cognitive skills described in our road map of human development (von Hofsten et al., in press), we believe the iCub to be a milestone in cognitive systems research by providing a solid framework for the community at large and for the first time providing opportunities for widespread collaborative progress. This is possible because of the opportunity of creating critical mass, using a common robotic platform and common software architecture, with the availability of technical support from an enthusiastic multidisciplinary team of developers, researchers and cognitive scientists. This places the iCub at the forefront of research in cognitive systems and robotics and fosters truly international collaboration by its adoption of the Open Source model.

Acknowledgement

This work was supported by the European Commission, Project IST-004370 RobotCub, under Strategic Objective 2.3.2.4: Cognitive Systems.

References

- Arbib, M. A. (2000). The mirror system, imitation, and the evolution of language. In C. L. Nehaniv, & K. Dautenhan (Eds.), *Imitation in animals and artifacts*. Cambridge, MA: The MIT Press.
- Borron, P., Montagna, M., Cerri, G., & Baldissera, F. (2005). Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement. *Brain Research* 1065, 115-124.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 1-5.

- Degallier, S., Righetti, L., Natale, L., Nori, F., Metta, G., & Ijspeert, A. (2008). A modular, bio-inspired architecture for movement generation for the infant-like robot iCub. In IEEE RAS/EMBS international conference on biomedical robotics and biomechanics (BioRob), Scottsdale, AZ, USA
- de Vries, J. I. P., Visser, G. H. A., & Prechtl, H. F. R. (1982). The emergence of fetal behaviour. I. Qualitative aspects. *Early Human Development*, 23, 159-191.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* 15, 399.
- Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology* 35(2-3), 165-177.
- Fitzpatrick, P., Metta, G., & Natale, L. (2008). Towards long-lived robot genes. *Journal of Robotics and Autonomous Systems* 56, 1-3.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308(4), 662-667.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting and knowing: toward an ecological psychology* (pp. 67-82). Lawrence Erlbaum.
- Huston, S. D., Johnson, J. C. E., & Sydd, U. (0000). *The ACE programmer's guide*. Addison-Wesley.
- Jeannerod, M. (1997). *The cognitive neuroscience of action* Cambridge, MA, Oxford, UK: Blackwell.
- Johnson, M. H. (1997). *Developmental cognitive neuroscience* Malden, MA, Oxford, UK: Blackwell.
- Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: the case of face recognition*. Oxford: Blackwell.
- Maggiali, M., Cannata, G., Maiolino, P., Metta, G., Randazzo, M., & Sandini, G. (2008). Embedded distributed capacitive tactile sensor. In *The 11th mechatronics forum biennial international conference*. University of Limerick, Ireland.
- Maturana, H. (1970). *Biology of cognition*. Research report BCL 9.0 University of Illinois, Urbana, Illinois.
- Maturana, H. (1975). The organization of the living: a theory of the living organization. *International Journal of Man Machine Studies*, 7(3), 313-332.
- Maturana, H. R., & Varela, F. J. (1980). *Boston studies on the philosophy of science Autopoiesis and cognition the realization of the living*. Dordrecht, Holland: D. Reidel Publishing Company.
- Maturana, H., & Varela, F. (1987). *The tree of knowledge the biological roots of human understanding*. Boston, London: New Science Library.
- Metta, G., Fitzpatrick, P., & Natale, L. (2006). Yarp: yet another robot platform. *International Journal of Advanced Robotic Systems* 3(1), 43-48.
- Metta, G., Sandini, G., & Konczak, J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12(10), 1413-1427.
- Metta, G., Sandini, G., Natale, L., & Panerai, F. (2001). Development and robotics. In IEEE RAS international conference on humanoid robots. Tokyo, Japan
- Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. In R. Madhavan, & E. Messina (Eds.), *Proceedings of the performance metrics for intelligent systems workshop (PerMIS)*: Vol. 1090 (pp. 50-56). NIST Special Publication.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford University Press.
- Montesano, L., Lopes, M., Bernardino, A., & Santos-Victor, J. (0000). Learning object affordances: from sensory motor maps to imitation. *IEEE Transactions on Robotics* 24 (1) [Special Issue on Bio-Robotics].
- Paolo, E. D., Rohde, M., & Jaegher, H. D. (2008). Horizons for the enactive mind: values, social interaction, and play. In J. Stewart, O. Gapenne, & E. D. Paolo (Eds.), *Enaction: towards a new paradigm for cognitive science*. Cambridge, MA: MIT Press.
- Pattacini, U., Nori, F., Natale, L., Metta, G., & Sandini, G. (2010). An experimental evaluation of a novel minimum-jerk Cartesian controller for humanoid robots. In 2010 IEEE/RSJ international conference on intelligent robots and systems. Taipei, Taiwan (in press).
- Perrett, D. I., Mistlin, A. J., Harries, M. H., & Chitty, A. J. (1990). Understanding the visual appearance and consequence of hand action. In M. A. Goodale (Ed.), *Vision and action: the control of grasping* (pp. 163-180). Norwood, NJ: Ablex.
- Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences* 20, 537-596.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences* 21(5), 188-194.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3, 131-141.
- Ruesch, J., Lopes, M., Hornstein, J., Santos-Victor, J., & Pfeifer, R. (2008). Multimodal saliency-based bottom up attention a framework for the humanoid robot iCub. In *International conference on robotics and automation*. Pasadena, CA, USA (pp. 962-967).
- Sakata, H., Taira, M., Kusunoki, M., Murata, A., & Tanaka, Y. (1997). The tins lecture the parietal association cortex in depth perception and visual control of action. *Trends in Neurosciences* 20(8), 350-358.
- Sandini, G., Metta, G., & Konczak, J. (1997). *Human sensori-motor development and artificial systems*.
- Shanahan, M. P. (2005a). Emotion, and imagination: a brain-inspired architecture for cognitive robotics. In *Proceedings AISB 2005 symposium on next generation approaches to machine consciousness* (pp. 26-35).
- Shanahan, M. P. (2005b). Cognition, action selection, and inner rehearsal. In *Proceedings IJCAI workshop on modelling natural action selection* (pp. 92-99).
- Shanahan, M. P. (2006). A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* 15, 433-449.
- Spelke, E. S. (2000). Core knowledge. In *American Psychologist* (pp. 1233-1243).
- Thompson, E. (2007). *Mind in life: biology, phenomenology, and the sciences of mind*. Boston: Harvard University Press.
- Tsagarakis, N. G., Metta, G., Sandini, G., Vernon, D., Beira, R., & Santos-Victor, J. et al. (0000). iCub the design and realisation of an open humanoid platform for cognitive and neuroscience research. *International Journal of Advanced Robotics* 21 (10), 1151-1175.
- Ungerleider, L. G., & Mishkin, M. (1982). Two visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.
- Varela, F. (1979). *Principles of biological autonomy*. New York: Elsevier North Holland.
- Varela, F. J. (1992). Whence perceptual meaning? A cartography of current ideas. In F. J. Varela, & J.-P. Dupuy (Eds.), *Boston studies in the philosophy of science Understanding origins contemporary views on the origin of life, mind and society* (pp. 235-263). Dordrecht: Kluwer Academic Publishers.
- Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation* 11(2), 151-180.
- von der Malsburg, C., & Singer, W. (1988). Principles of cortical network organisations. In P. Rakic, & W. Singer (Eds.), *Neurobiology of the neocortex* (pp. 69-99). London: John Wiley & Sons Ltd.
- von Hofsten, C. (2003). On the development of perception and action. In J. Valsiner, & K. J. Connolly (Eds.), *Handbook of developmental psychology* (pp. 114-140). London: Sage.
- von Hofsten, C. (2004). An action perspective on motor development. *Trends in Cognitive Sciences* 8, 266-272.
- von Hofsten, C., Fadiga, L., & Vernon, D. (2010). A roadmap for the development of cognitive capabilities in humanoid robots. In *COSMOSE* Elsevier (in press).