Cañamero, L., Oudeyer, P.-Y., and Balkenius, C. (Eds) Proceedings of the Ninth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems. Lund University Cognitive Studies, 146.

A Cognitive Robotic Model of Grasping

Zoran Macura* **Angelo Cangelosi*** **Rob Ellis**** **Davi Bugmann****
Martin H. Fischer*** **Andriy Myachykov*****
*School of Computing & Mathematics
**School of Psychology
University of Plymouth
Plymouth, PL4 8AA UK
{zoran.macura,a.cangelosi,r.ellis,d.bugmann}@plymouth.ac.uk

***Department of Psychology
University of Dundee
Dundee, DD1 4HN UK
{m.h.fischer,a.myachykov}@dundee.ac.uk

Abstract

In this paper we present a cognitive robotic model of object manipulation (i.e. grasping) based on psychologically plausible embodied cognition principles. Specifically, the robotic simulation model is inspired by recent theories of embodied cognition, in which vision, action and semantic systems are linked together in a dynamic and mutually interactive manner. The robotic agent is based on a simulation model of the iCub humanoid robot. It uses a connectionist control system trained with experimental data on object manipulation. Simulation analyses show that the robot is capable to reproduce phenomena observed in human experiments, such as the Stimulus-Response Compatibility effect.

1. Introduction

The primary aim of our work is to develop a cognitive robotic model of the processes involved in object grasping and manipulation following the embodied cognition view of action and vision integration and micro-affordance effects (Tucker and Ellis, 2001). The task typically involves how to select, based on the agent's knowledge and representations of the world, one object from several, grasp the object and use it in an appropriate manner. This mundane activity in fact requires the simultaneous solution of several deep problems at various levels. The agent's visual system must represent potential target objects, the target must be selected based on task instructions or the agent's knowledge of the functions of the represented objects, and the hand (in this case) must be moved to the target and shaped so as to grip it in a manner appropriate for its use.

This work will first be framed within the current literature on the psychological investigation on action, vision and language integration, and on the robotics and computational models of these cognitive phenomena. We will then present a simulation model of grasping based on the iCub humanoid platform. We discuss how this will be extended to perform experiments replicating known psychological data on micro-affordance effects and action/vision integration.

1.1 Psychological Studies on Vision, Action and Language

It is increasingly recognised that cognition should not be regarded as a set of disembodied processes, but is strongly determined by the constraints of its bodily implementation and it being situated in the world with which it interacts. In the case of visual cognition this embodied approach has led to an emphasis on the role of active vision in exploring the world, and therefore on the integration of vision and action (see for instance O'Regan and Noe, 2001). There is certainly accumulating human behavioural evidence that vision and action form a closely integrated and highly dynamic system (e.g. Tucker and Ellis, 1998, 2001; Craighero et al., 2002; Fischer and Hoellen, 2004).

One consequence of this integration of the vision and action systems is that seeing an object, even when there is no intention to handle it, potentiates elements of the actions needed to reach and grasp it. For instance participants who viewed photographs of common objects in order to decide whether they were manufactured or organic were facilitated in responding if the grip needed to make the response was one that could be used to handle the viewed object (Tucker and Ellis, 2001). So, for example, signalling that a pea was organic was easier (faster and more accurate) if a precision grip (using only the thumb and forefinger) was needed for the response compared to using a power grip (between the four fingers and palm). Similar object to action compatibility effects are observed for the hand of reach and the wrist rotation required to align the hand with an object (Tucker and Ellis, 1998; Ellis and Tucker, 2000). The authors coined the term 'micro-affordances' to describe these potentiated elements of an action.

Visual attention and eye movements are obviously fundamental components of human exploratory behaviour, and implicated in the integration of vision, action and language. Our eyes are exquisitely sensitive to the combined demands of vision, action and language processing. We move our eyes to project objects of interest onto the foveal area of high visual resolution. When we interact with objects, our eyes move ahead of the hand to support the on-line control of grasping (e.g. Bekkering and Neggers, 2002). Merely seeing objects activates plans for actions directed to them (e.g. Tucker and Ellis, 2001; Fischer and Dahl, 2007).

1.2 Computational Modelling of Vision, Action and Language

Researchers from different fields such as engineering and cognitive science, to name a few, have greatly benefited from the use of computational models. This has resulted in a phletora of computational approaches, amongst which some are based on cognitive and developmental robotics approaches. Such approaches provide us with a more integrative vision of action, language and cognition.

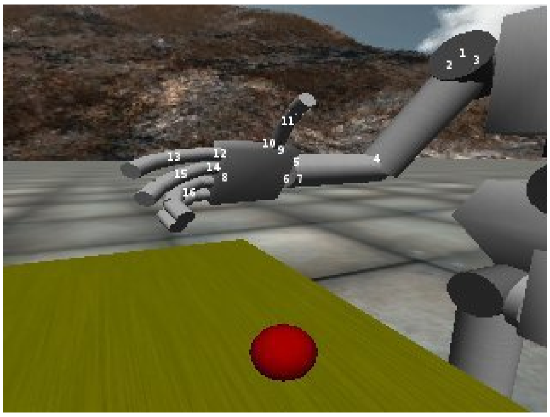
In the cognitive modelling literature, there has also been some work specifically focused on the integration of action and vision knowledge in cognitive agents and in connectionist models. For example, Arbib and colleagues have developed a neural model for action learning directly inspired by brain imaging studies on grasping in primates, and applied to action imitation learning simulations (Arbib et al., 2000). Haruno et al. (2001) proposed the Mosaic architecture for simulated object manipulation tasks, demonstrating that the model can generalise action-object associations depending on the object shape. Demiris and Simmons (2006) present a computational arctitecture using a hierarchical controller based on the minimum variance model of movement control (HAMMER: Hierarchical Attentive Multiple Models for Execution and Recognition) for implementing biologically plausible human reaching and grasping movements. Tsiotas et al. (2005) developed an artificial life model for simulating some of Tucker and Ellis (2001) findings. They used a simplified 2D arm model to study the evolutionary learning of object micro-affordances.[1] In the area of connectionist modelling, Yoon et al. (2002) have proposed a neural network model for action and name selection for objects (NAM: Naming and Action Model) that supports the role of a direct perception-action route for action selection. This model uses abstract (localist) encoding of action, perceptual and semantic information, rather than providing a robotic implementation, but is useful as it focuses on the comparison of perceptual vs semantic information in action selection.

More recently, Caligiore et al. (2008) developed a biomimetic neural network constrained by anatomical, physiological and behavioural data in which an embodied 'eye-hand' system was used to interact with objects of varying sizes (i.e. small and large). Using this model they replicated Tucker and Ellis (2001) compatibility effect between object size and the type of grip used in a categorisation task on whether objects were natural or artefacts. The modules of this neural network system are directly inspired by known brain processing mechanism. The action properties of the agents behaviour are however limited to a static representation of the action representing the final grasping configuration.

Models of action and vision integration also provide a framework to develop models of language learning based on the symbol grounding approach (Harnad, 1990; Cangelosi et al., 2005). Numerous studies have recently focused on the design of linguistic communication between autonomous agents, such as robots or simulated agents. The agents' linguistic abilities in these models are strictly dependent on, and grounded in, other behaviours and skills such as vision and action. Numerous sensorimotor, cognitive, neural, social and evolutionary factors contribute to the emergence and establishment of communication and language. For example, in these models there exists an intrinsic link between the communication symbols (words) used by the agent and its own cognitive representations (meanings) of the perceptual and sensorimotor interaction with the external world (referents), as denoted by these symbols. Such a grounded and embodied approach to language design is consistent with the psychologically-plausible theories of the grounding of language (Cangelosi and Riga, 2006).

In such cognitive robotic models, communication results from the dynamical interaction between the robot's physical body, its cognitive system and the external physical and social environment. Some studies stress the grounding in action and sensorimotor processes, such as Marocco et al.'s (2003) model of robotic arms and Vogt's (2001) mobile

[1]Recall that a micro-affordance is a quality of an object which is perceivable by an individual and suggests to this individual a range of possible actions associated with it.



#	joint	minimum angle (degrees)	maximum angle (degrees)
1	shoulder pitch	-95	90
2	shoulder roll	0	161
3	shoulder yaw	-37	100
4	elbow	6	106
5	wrist pronosupination	-90	90
6	wrist pitch	-90	10
7	wrist yaw	-20	40
8	hand finger adduction/abduction	-20	30
9	thumb opposition	-15	105
10	thumb proximal flexion	0	90
11	thumb distal flexion	0	90
12	index proximal flexion	0	90
13	index distal flexion	0	90
14	middle proximal flexion	0	90
15	middle distal flexion	0	90
16	ring & little flexion	0	115

Figure 1: Simulated robot arm and hand with 16 controlled DoF and their corresponding movement ranges

robots. Other robotic models highlight the grounding through social interaction, such as Steels and Kaplan's (2001) AIBO robots. On the other hand, some studies are based on simulating adaptive agents. They model the agent and its environment with a good degree of detail upon which emergent meanings can be directly constructed. These simulation models have focused on grounding in perceptual experience and in cognitive representations and sensorimotor interactions (e.g. Cangelosi, 2001).

In the next section we present a preliminary robotic model of action and vision integration for a grasping task that is directly inspired by this experimental literature on embodiment. This model provides us with a test-bed for the simulation of the vision-action-language integration processes observed in psychology experiments, and generates further insights and prediction on such phenomena.

2. Model

The cognitive robotic model presented here is directly inspired by recent theories of embodied cognition, in which the vision, action and semantic systems are linked together, in a dynamic and mutually interactive manner, within a connectionist architecture. We take inspiration from the Caligiore et al. (2008) model described above and extend it to consider a more realistic simulation of grasping behaviour and its time dynamics. This model proposes a combination of the epigenetic robotics methodologies with the "embodied connectionist" modelling approach. Epigenetic (developmental) robotics is based on the use of embodied robotic systems that are situated in a physical and social environment and are subject to a prolonged epigenetic developmental process for the acquisition of cognitive capabilities (Weng et al., 2001; Lungarella et al., 2003; Schlesinger et al., 2008). Embodied connectionism refers to the use of artificial neural networks for the learning and control of behaviour in cognitive robotic agents. The integration of robotics and connectionist methodologies permits the transfer of the principles and advantages of connectionism and parallel distributed processing systems into embodied robotic agents (Cangelosi and Riga, 2006).

2.1 Simulated Robot

The robotic agent used in the simulation experiments is based on the humanoid iCub robot (Metta et al., 2008). In particular, the experiments use the recently developed open-source simulator of the iCub robot (Tikhanoff et al., 2008). The simulator has been designed to reproduce, as accurately as possible, the physics and the dynamics of iCub robot and its environment. The simulated iCub robot is composed of multiple rigid bodies connected via joint structures. It has been constructed collecting data directly from the robot design specifications in order to achieve an exact replication (e.g. height, mass, Degrees of Freedom) of the first iCub prototype developed at the Italian Institute of Technology in Genoa. The environment parameters on gravity, objects mass, friction and joints are based on known environment conditions.

The iCub robot is around 105cm high, weighs approximately 20.3kg and has a total of 53 degrees of freedom (DoF). These include 12 controlled DoF for the legs, three controlled DoF for the torso, six for the head and 32 for the arms. In particular, each arm is made up of three components (the arm, the forearm and the hand) where the arm and forearm include eight DoF and the hand another eight DoF, where each DoF movement range is constrained with the respective human DoF movement range (Fig. 1).

The robot's vision system consists of two cameras located at the eyes of the robot. The simulated robot also has touch and force/torque sensors which receive tactile information and proprioceptive data on its own body posture. The proprioceptive sensors are located on the robot's arm and hand and encode the

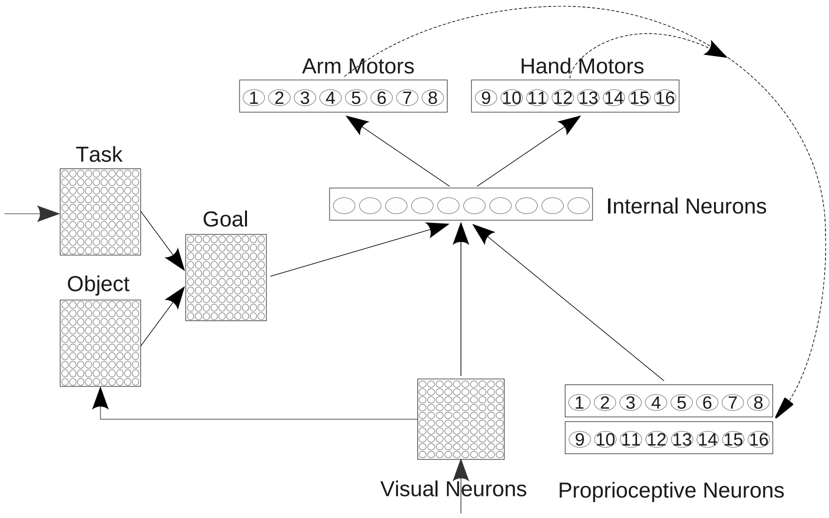


Figure 2: Neural network architecture for the robotic agent in which according to the visual input, the proprioceptive data and the task instruction an appropriate grasping movement is made

current angles of all 16 DoF of the arm (as listed in Figure 1). In addition, there are six tactile sensors in each hand – one on each finger and one on the palm – which indicate whether that particular body part is in physical contact with another object.

The simulator has full interaction with the world/environment. The objects within this world can be dynamically created, modified and queried which enables us to train the robotic agent to interact with objects so that it can acquire a sensorimotor representation of the objects through eye and hand movements. In learning to act on objects the robots neural controller will form embodied representations of those objects and as a consequence future encounters with these objects will cause them to afford the associated actions (micro-affordances).

2.2 Neural Network Architecture

A connectionist network is used to learn and guide the behaviour of the robot and to acquire embodied representations of objects and actions. The neural architecture, based on the Jordan recurrent architecture, has recurrent connections to permit information integration and the execution of actions such as grasping (Marocco et al., 2003). The network is depicted in Figure 2, and it is made up of four 2D maps of 10x10 neurons, 16 proprioceptive neurons, 10 internal neurons and 16 motor neurons. Specifically, the 16 output motor neurons control the DoF of the robot's right arm (see Fig. 1) that performs the grasping task and the 16 proprioceptive neurons in input encode the current angles of the right arm's DoF, which feed into the neural network thus creating a recursive structure (Fig. 2).

The visual input to the robot's neural controller consists of pre-processed information regarding visual object properties (i.e. shape and size). This information is processed directly from the physics simulator (Fig. 3: *Real Image*) by using three edge-detection Sobel filters – where each filter is sensitive to either red, green or blue component of the object's colour. The result of the Sobel filtering is an image where only the edges of the objects in vision are encoded (Fig. 3: *Sobel*). An assumption in this model is that the eyes always foveate the target object. The foveated area of the image is in turn processed – where the activated edges are encoded as 1 and everything else as 0 – resulting in a 10x10 2D map encoding the shape and size of the foveated object (Fig 3: *Visual Input*), which constitutes the visual input for the neural network.

As well as being fed into the internal neurons, the visual input is also fed into the *Object* map, which is used to encode the objects' identity. On the other hand, the *Task* map encodes the different tasks that can be performed on objects, namely a normal grasping task or a categorisation task akin to Tucker and Ellis (2001) psychological experiments. *Object* and *Task* maps in turn feed into the *Goal* map, which encodes the information about the current goal of action depending on the task and object identity. These three maps are implemented as Kohonen self-organising maps (SOMs) and are analogous to inferior temporal cortex (IT), medial temporal cortex (MT) and prefrontal cortex (PFC) in Caligiore et al.'s (2008) model, where the use of Kohonen maps for IT and PFC is justified by studies suggesting that these cortical areas are involved in high-level visual processing and categorisation (Miller et al., 2002; Shima et al., 2007). Note that these SOMs are just an approximation of the relevant cortices

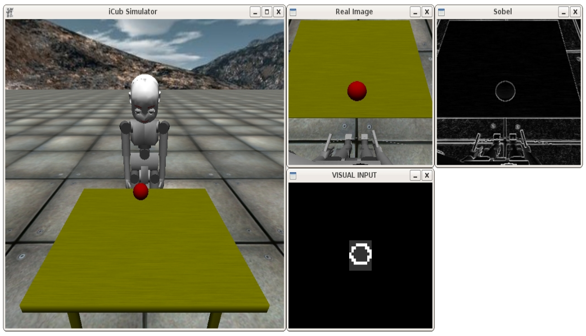


Figure 3: Visual processing in the robotic simulation. *Real Image* is the image taken from the robot's eyes. The eyes always foveate on the target object, after which Sobel filters are applied to the *Real Image* producing the *Sobel* image. The result of the visual processing is the *Visual Input* image – a 10x10 2D map of 0s and 1s encoding the shape and size of the foveated object

in functional terms, and not the real physiological-anatomic analogues.

2.3 Training

Simulation experiments focus on the training of the robot to use objects using different manipulation modalities (e.g. precision grip vs power grip, respectively, for small objects – "cherries" – and for big objects – "apples") and also to be able to replicate psychological experiments where the objects can be categorised using different grips (e.g. precision grip for artefacts and power grip for natural objects).

There are four objects in the simulation: 2 larger objects ('big-ball' and 'big-cube') for power grips; and 2 smaller objects ('small-ball' and 'small-cube') for precision grips. In this model, the round objects (big and small balls) are viewed as natural objects, whereas the cubes are viewed as artefacts. The training data consists of a set of grasping sequences for each object, which have been normalised in the range of 0–1 from the movement ranges shown in Figure 1. Each sequence is made up of 10 time-based steps where the first step represents the initial arm and hand position (pre-grasp) whereas the last step represents the final grasping posture (appropriate grasp for the target object).

There are two training phases in the model. In the first training phase the robot learns to appropriately grasp objects, while in the second phase the robot learns how to categorise objects using power and precision grips, as seen in psychological experiments. Before the main training begins, the *Object*, *Task* and *Goal* SOMs are trained individually offline. The *Object* SOM is trained to categorise the four objects where a different cluster of neurons is activated for each oject, and the size of the cluster is dependant on the object size (e.g. large objects activate a greater cluster of neurons) (see Caligiore et al., 2008). The *Task* SOM is trained to activate two different patterns of neurons to represent the two different tasks in psychological experiments, namely normal grasp and categorisation task. Finally, the *Goal* SOM is trained to represent the current goal, where there are eight different clusters of neurons that can get activated depending on the object and the task.

During the first training phase, the four objects are repeatedly presented to the simulated robot, which in turn tries to learn the micro-affordance-based behaviours for each object. At the beginning of each trial an object is placed at the same position on the table and the robot foveates the object. The processed visual input is then fed into the neural network along with the proprioceptive data encoding the current position of the robot's right arm DoF. The *Task* SOM in this phase is always activated with the pattern representing the grasping task. Each internal neuron performs a weighted summation of the inputs, which then passes a sigmoid (nonlinear) activation function. The motor neurons perform a similar weighted summation of the internal neurons' outputs and their outputs result in a grasping movement appropriate for the target object. In this learning phase the network parameters are continuously adjusted using a back-propagation algorithm until the robot learns to form appropriate associations between the object's shape and the hand shape.

In the second training phase, the robot learns to categorise objects with different grips. The training follows a similar procedure to the one outlined above, with the main difference being in the way the *Task* SOM is activated. In this phase, the *Task* SOM instead of always being activated with the pattern representing the grasping task (as was the case in the first training phase), is first activated with a new (random) pattern representing the categorisation tasks, and in the next cycle with the previous pattern of the grasping task. This enables the robot to learn suitable grasps to correctly categorise objects depending on the *Object* and *Task* SOMs activations.

3. Results

A total of five different grasping sequences were defined for each object – of which four sequences were used for training the robot and the fifth was used for testing purposes. The five grasping postures differ by the final position and rotation of the hand with respect to the object. The learning rate for the back-propagation algorithm was set to 0.075. For each training cycle an object and one of its four grasping sequences was chosen randomly and presented to the robot. This was repeated 12000 times, where

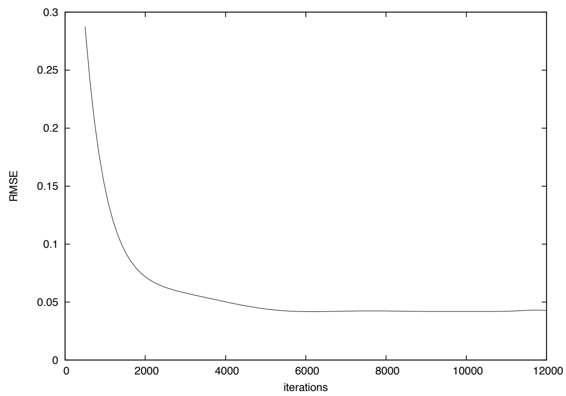


Figure 4: Average RMSE during the neural network training

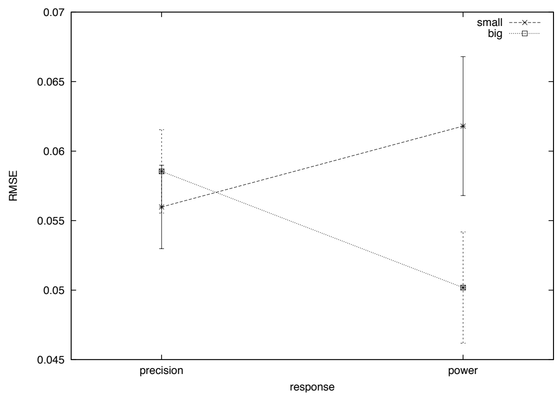


Figure 5: Compatibility effect in congruent and incongruent trials

both training phases lasted 6000 iterations each.[2] After training, the network was tested on both grasping and categorisation tasks for all five grasping sequences – including the fifth (unseen) grasping sequence – for each of the four objects in order to establish whether the robot learned how to grasp and categorise objects appropriately. The results presented here are all averages of 12 trained neural networks, where the first six networks were trained to categorise natural objects (balls) with a power grip and artefacts (cubes) with a precision grip and the other six networks were trained to do the opposite (power grip for artefacts and precision grip for natural objects).

Figure 4 shows the root mean squared error (RMSE) of the network during the training phase. As expected, at the begining of the training the error between the motor outputs and the desired targets (joint positions of the right arm) is high (around 0.3). After roughly 2000 iterations the RMSE drops to 0.05 and stabilises around this value, indicating that the simulated robot has been able to successfully learn appropriate grasps for the four objects.

One important test in this model of object grasping and micro-affordances is the comparison of the congruent (where the categorisation grip is in agreement with the natural grip) and incongruent (where there is mismatch between the categorisation grip and the natural grip) conditions. The trained neural networks were presented each object in turn, where the desired target depended on the task being performed. The results are depicted in Figure 5, which shows the average RMSE values of the 12 networks for congruent and incongruent trials. We assume that RMSE is analogous to reaction time used in psychological experiments done in the Tucker and Ellis (2001) study. An ANOVA on response times was performed with two factors: congruency and object size, and both factors were statistically significant. As can be seen in Figure 5 the results are in agreement with psychological experiments where reaction times are faster in congruent than in incongruent trials. In addition, the reaction times for larger objects were faster than for smaller object, as was also the case in psychological experiments. This indicates that the robot was able to generalise a grasping sequence for each task and object from the four grasping sequences used in training, hence learning to appropriately grasp and categorise objects based on their shapes and sizes.

4. Conclusion & Future Work

We have shown how the proposed cognitive robotic model was able to learn object micro-affordances and appropriately grasp and categorise an object depending on its shape and size. Tests on congruent/incongruent tasks also demonstrate that the robots neural controller uses micro-affordance information about the objects replicating the well known Stimulus-Response Compatibility phenomenon observed in psychology experiments. Future analyses will investigate the internal representations used by the network in responses to various task demands (grasping vs. categorisation), to different level of object/grasp congruency and to the interaction between objects with conflicting micro-affordance. Analyses of the neural network representations in controlling behaviour, and of the time-course of processes and representation activated by the robot's neural controller, will be used to better understand behaviour observed in human participants and to derive novel predictions about interactions between vision and action.

One additional extension of this model regards the

[2] Recall that in the first phase the *Task* SOM is always activated with the pattern representing the grasping task. In the second phase this happens only in half of the cases (3000 iterations) and in the other half the *Task* SOM is activated with the pattern representing the categorisation task.

inclusion of linguistic information during training, such as for the names of objects and actions. This extended model will permit detailed investigations of the effects of language on micro-affordance effect (Tucker and Ellis, 2004).

The main goal of this psychologically-plausible model for the study of grasping behaviour in humanoid robots, in addition to advancing our understanding of vision-action-language integration, will provide us with a set of cognitively-plausible design principles for developing vision, action and linguistic capabilities in robots and their use in interactive cognitive systems and autonomous robotics.

The cognitive robotic platform developed here can be used as a tool to test feasibility of the vision-action-language integration mechanisms identified during experimental studies, in addition to demonstrating the technological potential in such an approach. Observation and analyses of the robot's cognitive and linguistic capabilities will also result in the production and test of new predictions about mechanisms integrating vision, action and language. The replication in a robotic model of the psychological phenomena observed in experimental studies will have the advantage of permitting the fine analysis and understanding of the neural and behavioural processes that contribute to action-vision-language integration (Cangelosi and Parisi, 2002).

Acknowledgements

This research is supported by the VALUE project (EPSRC Grant EP/F026471) and the iTalk project (EU FP7).

References

Arbib, M. A., Billard, A., Iacoboni, M., and Oztop, E. (2000). Synthetic brain imaging: Grasping, mirror neurons and imitation. *Neural Networks*, 13:975–997.

Bekkering, H. and Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychological Science*, 13:370–374.

Caligiore, D., Borghi, A. M., Parisi, D., and Baldassarre, G. (2008). Affordances and compatibility effects: a neural-network computational model. In *The 11th Neural Computation and Psychology Workshop*. University of Oxford, Oxford, UK.

Cangelosi, A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*, 5(2):93–101.

Cangelosi, A., Bugmann, G., and Borisyuk, R., (Eds.) (2005). *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. World Scientific, Singapore.

Cangelosi, A. and Parisi, D., (Eds.) (2002). *Simulating the evolution of language*. Springer-Verlag, London.

Cangelosi, A. and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30(4):673–689.

Craighero, L., Bello, A., Fadiga, L., and Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, 40:492–502.

Demiris, Y. and Simmons, G. (2006). Perceiving the unusual: temporal properties of hierarchical motor representations for action perception. *Neural Networks*, 19(3):272–284.

Ellis, R. and Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, 91:451–471.

Fischer, M. H. and Dahl, C. (2007). The time course of visuo-motor affordances. *Experimental Brain Research*, 176(3):519–524.

Fischer, M. H. and Hoellen, N. (2004). Space-based and object-based attention depend on motor intention. *Journal of General Psychology*, 131(4):365–377.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42:335–346.

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning. *Neural Computation*, 13:2201–2220.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: A survey. *Connection Science*, 15(4):151–190.

Marocco, D., Cangelosi, A., and Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 361(1811):2397–2421.

Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). The icub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems (PerMIS08)*.

Miller, E. K., Freedman, D. J., and Wallis, J. D. (2002). The prefrontal cortex: categories, concepts, and cognition. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 357:1123–1136.

O'Regan, J. K. and Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Brain and Behavioral Sciences*, 24(5):939–1031.

Schlesinger, M., Berthouze, L., and Balkenius, C., (Eds.) (2008). *Proceedings of the Eighth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems.* Lund University Cognitive Studies, 139, LUND: LUCS.

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature*, 445:315–318.

Steels, L. and Kaplan, F. (2001). Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator. In *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems (PerMIS08)*.

Tsiotas, G., Borghi, A., and Parisi, D. (2005). Objects and affordances: An artificial life simulation. In *Proceedings of the Cognitive Science Society*, pages 2212–2217.

Tucker, M. and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24:830–846.

Tucker, M. and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8(6):769–800.

Tucker, M. and Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, 116:185–203.

Vogt, P. (2001). Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, 4(1):87–116.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291:599–600.

Yoon, E. Y., Heinke, D., and Humphreys, G. W. (2002). Modelling direct perceptual constraints on action selection: The naming and action model. *Visual Cognition*, 9(4/5):615–661.