A Neural Network model for spatial mental imagery investigation: A study with the humanoid robot platform iCub

Alessandro G. Di Nuovo and Davide Marocco and Santo Di Nuovo and Angelo Cangelosi

***Abstract*—Understanding the process behind the human ability of creating mental images of events and experiences is a still crucial issue for psychologists. Mental imagery may be considered a multimodal biological simulation that activates the same, or very similar, sensorial and motor modalities that are activated when we interact with the environment in real time. Neuro-psychological studies show that neural mechanisms underlying real-time visual perception and mental visualization are the same when a task is mentally recalled. Nevertheless, the neural mechanisms involved in the active elaboration of mental images might be different from those involved in passive elaborations. The enhancement of this active and creative imagery is the aim of most psychological and educational processes, although, more empirical effort is needed in order to understand the mechanisms and the role of active mental imagery in human cognition. In this work we present some results of on ongoing investigation about mental imagery using cognitive robotics. Here we focus on the capability to estimate, from proprioceptive and visual information, the position into a soccer field when the robot acquires the goal. Results of simulation with the iCub platform are given to show that the computational model is able to efficiently estimate the robot's position. The final objective of our work is to replicate with a cognitive robotics model the mental imagery when it is used during the training phase of athletes that are allowed to imaginary practice to score a goal.**

I. Introduction

Robotics, as well as many other information technology advancements, allows scientists interested in artificial cognitive systems to 'create' autonomous or partially autonomous robots capable to deal with a great number of problems that arise in the real-world scenarios where many robots operates nowadays (especially in factories and mass-production contexts). Nevertheless, many challenges still exist. In particular, the new and fast growing field of humanoid robotics, despite the tremendous potentiality of future applications, still poses several interesting challenges, both in terms of mechanics and autonomous control. Improving the skills of a humanoid robot for complex sensorimotor tasks is still regarded as a complex problem in current robotics research. In humanoid robots, in particular, sensors and actuators arrangements determine a highly redundant morphological structure, which is traditionally difficult to control. Interestingly, when Artificial Intelligence was moving it first steps, Alan Turing argued that in order to think and speak a machine may need a human-like body and that the development of robot cognitive skills might be just as simple as teaching a child. This principle is reinforced by the concept of embodied cognition (e.g. [1], [2]), which affirms that the nature of intelligence is largely determined by the form of the body. Therefore, the body and every physical experience made through the body, shape the form of intelligence that can be observed in any autonomous systems. From this perspective, understanding the role of the body in cognitive processes is extremely important and psychological and neuroscience studies are extremely important in this regard. The psychologist Margaret Wilson, in [3], identified six claims in the current view of embodied cognition: (1) cognition is situated; (2) cognition is time-pressured; (3) we off-load cognitive work onto the environment; (4) the environment is part of the cognitive system; (5) cognition is for action; (6) offline cognition is bodily based. Among those six claims, the last claim is particularly important. According to this claim, sensorimotor functions that evolved for action and perception are used during offline cognition that occurs when the perceiver represents social objects, situations or events in times and places other than the ordinary ones. That means that even if the mind does not directly interact with the environment, it is able to apply mechanisms of sensory processing and motor control by using some innate abilities such as memory (implicit, short and long term), problem solving and mental imagery. These capabilities have been well studied in psychology and neuroscience, but the debate is still open on the issue of mental imagery, where mental imagery is defined as a sensation activated without sensorial stimulation. Among the many hypothesis and models already tested in the field of robot control, the use of mental imagery as a cognitive tools capable to enhance robots performance is both innovative and well grounded on experimental data. Recent psychological findings, both in experimental research and practical contexts, demonstrate that mental imagery is now applied to increase performance in professional and amateur sports and in motor rehabilitation. In cases of injury, for instance, mental simulation offers a way of training even when active movement execution is severely compromised. As a result, new opportunities for the use of mental training have opened up in the fields of medical and orthopedic-traumatologic rehabilitation. Mental training can be applied successfully in actively helping a person to regain lost movement patterns after joint operations or joint replacements and in neurological rehabilitation for after-stroke patients. Overall, modern psychological research clearly demonstrates the tight connection between mental training and motor performance improvement. The primary role of mental imagery

Davide Marocco and Angelo Cangelosi are with University of Plymouth (emails: {davide.marocco,a.cangelosi}@plymouth.ac.uk), Alessandro G. Di Nuovo and Santo Di Nuovo are with Universita´ degli Studi di Catania (emails adinuovo@ieee.org ; s.dinuovo@unict.it). This work was partially supported by the European Commission FP7 Project ITALK (ICT-214668) within the Cognitive Systems and Robotics unit (FP7 ICT Challenge 2).

in motor performance and in supporting human high-level cognitive capabilities involved in complex motor control and events anticipation. Many evidences from empirical sciences have demonstrated the relationship between bodily experiences and mental processes that involve body representation. Neuropsychological research has demonstrated that the same brain areas are activated during seeing or recalling by images [4] and that areas controlling perception are needed also for maintaining mental images active in working memory. Therefore, mental imagery may be considered as a kind of biological simulation. Jeannerod, in [5] observed that the primary motor cortex M1 is activated during the production of motor images as well as during the production of active movement. Similarly, the imagined time for a walk is proportional to the real action [6]. These studies demonstrate the tight relationship between mental imagery and motor activities (i.e. how the image in mind can influence movements and motor skills). Psychologically speaking, mental imagery is defined as an internal experience that resembles a real perceptual experience and that occurs in the absence of appropriate external stimuli. The role of mental imagery has been researched extensively over the past 50 years in areas of motor learning and psychology. Nevertheless, although effects of mental practice on physical performances have been well established, processes involved and explanations are still largely unclear. Therefore, understanding the process behind the human ability of creating mental images of events and experiences is still a crucial issue. For this reason more empirical effort is needed in order to understand the mechanisms and the role of active mental imagery in human cognition.

Understanding the relation between sensory-motor skills and mental imagery is particular important in domains in which improving those skills is crucial for reaching better performances, such as sports and rehabilitation. In sport, for example, [7] demonstrated that sport experts showed more focused activation patterns in prefrontal areas while performing imagery tasks than novices. This may be relevant to higher order of motor control in motor imagery and it allows saying that the brains of sport experts could be considered as the ideal subjects to explore the relationship between cerebral plasticity and learning of complex motor skills. From a technological point of view, a better understanding of mental imagery in humans will give the opportunity of deriving engineering principles for the development of artificial cognitive systems capable to better interact with the environment and refine their motor skill in an open-ended process. In this work we present some preliminary results of on ongoing investigation about mental imagery using cognitive robotics. Here we focus on the capability to estimate, from proprioceptive and visual information, the position into a soccer field when the robot acquires the goal. The final objective of our work is to replicate with a cognitive robotics model the study presented in [8], where mental imagery is used during the training phase of athletes that are allowed to imaginary practice to score a goal. Results in [8] show that the imaginary practice enhances the train phase, obtaining better performance.

The rest of the paper is organized as follows: Section II presents the Materials and the Methods used, while Section III shows the results obtained in our experiments. Finally Section IV reports our conclusion and intention for future work.

II. Material and Methods

The robotic model used for the experiments is a simulation of the iCub humanoid robot controlled by a recurrent artificial neural network. The iCub platform is a child-like humanoid robot (approximately 1m tall), with 53 degrees of freedom distributed on the head, arms, hands and legs. It has been specifically designed to maximize the number of degrees of freedom allocated to the hands (with the constraint of the overall small size). An implementation of the iCub platform is in Figure 1. Experiments were conducted using an open-source iCub simulator[9]. The simulator allows creating realistic physical scenarios in which the robot can interact with a virtual environment. Physical constrains and interactions that occur between the objects of the environment are simulated using specific types of physics dynamics libraries that provide an accurate simulation of rigid bodies dynamics and collisions. The iCub simulator is currently implemented using ODE or Newton Game Dynamics. The simulator reflects the overall default body and actuator structure and DoFs of the iCub robot as described above. For the present work have been used the vision system and only a small subset of the available degrees of freedom. The environment is a square portion of a soccer field, whose length and width are both 15 meters. At one end is placed a goal 1.94m wide, that, as can be seen in Figure 3, is represented all in blue to contrast with the background and to be easily recognizable. The robot can be positioned anywhere in this square and the ball is placed in front of his left foot, as start position.

The neural system that controls the robot is a fully connected recurrent neural network (Figure 2) with 16 hidden units, 37 input units and 6 output units. Input variables are the following: visual information; polar coordinates of the robot with respect to the goal (represented by the radius and the polar angle); the current angle of the neck (left-right movement), the torso (left-right movement) and the body rotation joint (a joint attached to the body that allows the robot to rotate on its vertical axis). Visual information are provided through a vector of 32 bits, which encodes a simplified visual image obtained by the eyes and allows the robot to locate the goal in its visual field. The 6 outputs are, respectively, the polar coordinates of the robot respect to the goal, the desired angle for the neck, the torso and the body rotation joint, and an additional binary output that makes the robot kick the ball when its value is 1. In the learning phase this output was set to 1 only when the motion ends and the robot must kick the ball to the goal. All input and output variables are normalized in $[0, 1]$. Activations of hidden and output units y_i are calculated at a discrete time, by passing

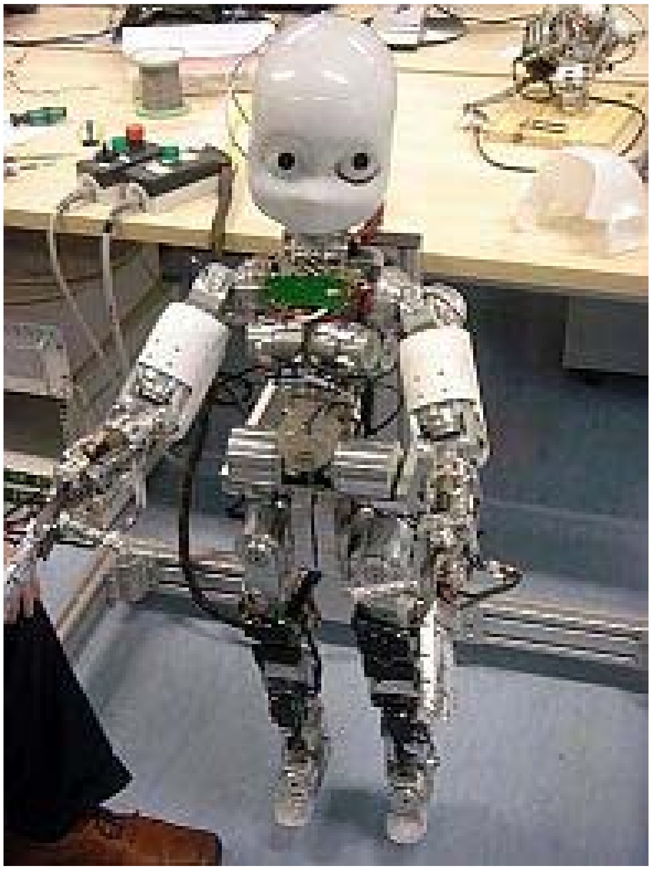

Fig. 1. An implementation of the iCub platform.

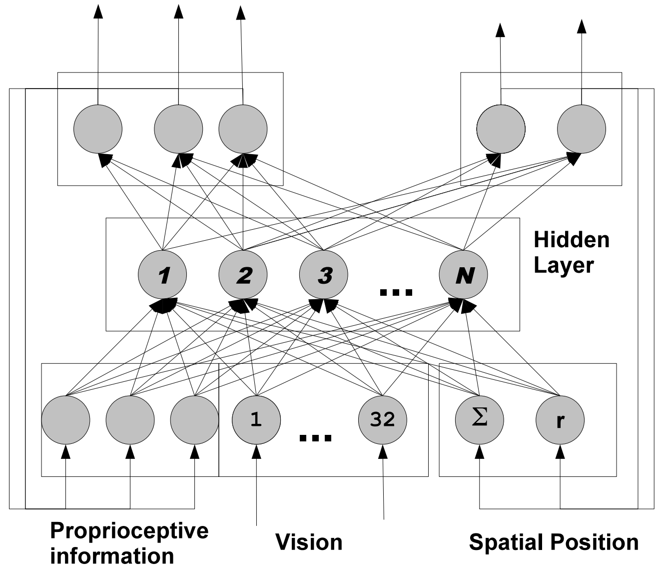

Fig. 2. Recurrent Neural Network

the net input u_i to the logistic function, as it is described in equations (1) and (2):

$$u_i = \sum_i y_i * w_{ij} - k_i \quad (1)$$

$$y_i = \frac{1}{1 - e^{-u_i}} \quad (2)$$

where w_{ij} is the synaptic weight that connects unit i with unit j and k_i is the bias of unit i. The output units encode the values of the input at the time step $t+1$. That is, the output state corresponds to the next input state of the network. The network is trained to predict its own input. As we will see in the next section, during the testing phase, the predicted input state is also used to provide target angles for the actuators.

The structure of the experiment is divided into two phases: in the first phase the network is trained to predict its own subsequent sensorimotor state. In the second phase the network is tested on the robot, in interaction with the environment.

For training the neural network we used the Back Propagation Through Time algorithm (BPTT), which is typically used to train neural network with recurrent nodes [10]. This algorithm allows a neural network to learn the dynamical sequences of input-output patterns as they develop in time. Since we are interested in the dynamic and time dependent processes of the robot-object interaction, an algorithm that allows to take into account dynamic events is more suitable than the standard Back-propagation algorithm [10]. For a detailed description of the BPTT algorithm see also [11]. The main difference between a standard Back Propagation algorithm and the BPTT is that, in the latter case the training set consists in a series of input-output sequences, rather than in a single input-output pattern. The BPTT allows the robot to learn sequences of actions. The goal of the learning process is to find optimal values of synaptic weights that minimize the error E, defined as the error between the teaching sequences and the output sequences produced by the network. The error function E is calculated as follows:

$$E = \sum_s \sum_t \sum_i ((y^*_{its} - y_{its})(y_{its} - (1 - y_{its})))^2 \quad (3)$$

where y^*_{its} is the desired activation value of the output unit i at time t for the sequence s and y_{its} is the actual activation of the same unit produced by the neural network, calculated using equation (1) and (2). During the training phase, synaptic weights at learning step $n+1$ are updated using the error δ_i calculated at the previous learning step n, that in turn depend on the error E (for details see [10]), according to the following equation (4):

$$\Delta w_{ij}(n+1) = \eta \delta_i y_j + \alpha \Delta w_{ij}(n) \quad (4)$$

where w_{ij} is the synaptic weight that connects unit i with unit j, y_i is the activation of unit j, α is the learning rate and η is the momentum.

For experiments we used the iCub simulator to position the robot into the environment in 8 different positions and then rotated to acquire the goal according to a simple searching algorithm. During this movement, proprioceptive and visual information were sampled in order to build 20 input-output sequences corresponding to the 8 positions. After this movement, the robot was allowed to kick the ball. The kicking movement was pre-programmed. With the 8 series recorded, the neural network was trained to predict the next sensory state (excluding the visual input) by means of a backpropagation through time algorithm [10] for 50.000 epochs, after which the MSE error in estimating the 6 output variables was 0.0056.

III. Results

Using the same material and methods, in a previous work ([12]) two case studies were carried out and confronted in the same scenario considered in this work. In a first case the robot was controlled by the same algorithm used for collecting the train series (*Controlled* condition), while the neural network was used to estimate the position coordinates only. In the second study the robot was fully moved by means of the neural network (*Autonomous* condition), which controlled the neck angle, the torso angle and the body rotation angle, as well as the output commanding the kick. Results shown that the *Autonomous* condition performs better than the *Controlled* one. In the *Controlled* condition the average error in estimating its position is the 14.125%, whilst in the *Autonomous* condition the error is 9.245%. In this work we studied only the *Autonomous* condition, testing the generalization performance of the neural network and evaluating the use of visual and proprioceptive information for the estimation of the robot position with respect to the goal. As testing phase, the robot was positioned in the same positions used for training (learning set) to verify the quality of the learning and, then, in new positions (testing set), that were not experienced before, to evaluate the generalization capability of the model. Results shown in this section refer to the testing phase only for both sets.

Table I details the learning and test positions, with their polar coordinates. Angle and Radius values are normalized in $[0, 1]$: Angle ranges from -90 degrees to 90 degrees, for this reason 0 degree (i.e. when the goal is in front of the robot) equals to 0.5; Radius could vary from 0 meters to 15 meters.

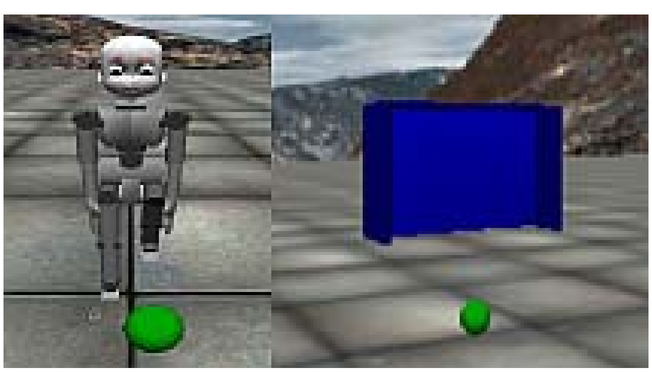

Fig. 3. iCub simulator: our test environment.

TABLE I
TRAIN AND TEST POSITIONS, POLAR COORDINATES, NORMALIZED VALUES

	Train		Test	
N	Angle	Radius	Angle	Radius
1	0.500	0.400	0.578	0.275
2	0.340	0.457	0.437	0.340
3	0.638	0.477	0.609	0.496
4	0.578	0.550	0.386	0.570
5	0.422	0.550	0.542	0.511
6	0.623	0.706	0.500	0.653
7	0.368	0.743	0.645	0.757
8	0.500	0.733	0.364	0.806

TABLE II
PERCENTAGE ERROR OF IMAGINED POSITIONS WITH RESPECT THE REAL ONES

N	Learning	Test
1	2.07	11.95
2	122.59	11.57
3	11.18	4.32
4	4.04	21.56
5	3.38	7.35
6	10.48	2.74
7	130.32	20.24
8	11.47	125.53
Avg	36.94	25.66
Avg without pos [2;7] / [8]	7.10	11.39

Table II shows the error in percentage between imagined positions and real positions. The error is evaluated as the percentage with respect to the real positions using the polar coordinates shown in Table I. All values of the imagined positions are in tables III and IV show results with *Autonomous* control. By analyzing the errors in testing positions, we observed that the error is very high for position 8 in which the robot fails to acquire the target because it makes a wrong move and the goal goes out of sight. It should be said that the robot has not been trained to find the goal when it is not in its visual field, at least in part. The same happened when the robot was in position 2 and 7 of the learning set. The robot misses the 50% of the scores, but it is worth to mention that errors were mostly made in the test set (5 out of 8) and when the goal was very far and even a little error in the position leads the ball far from the goal.

Figure 4 and 5 graphically summarize the results, showing

TABLE III
REAL AND IMAGINED VALUES FOR TRAIN POSITIONS

	Real values		Imagined Values	
N	Angle	Radius	Angle	Radius
1	0.500	0.400	0.503	0.415
2	0.340	0.457	0.764	0.905
3	0.638	0.477	0.640	0.560
4	0.578	0.550	0.561	0.545
5	0.422	0.550	0.408	0.543
6	0.623	0.706	0.657	0.724
7	0.368	0.743	0.764	0.905
8	0.500	0.733	0.504	0.648

TABLE IV
REAL AND IMAGINED VALUES FOR TEST POSITIONS

N	Real values Angle	Real values Radius	Imagined Values Angle	Imagined Values Radius
1	0.578	0.275	0.574	0.242
2	0.437	0.340	0.433	0.301
3	0.609	0.496	0.605	0.476
4	0.386	0.570	0.386	0.447
5	0.542	0.511	0.533	0.476
6	0.500	0.653	0.509	0.652
7	0.645	0.757	0.582	0.780
8	0.364	0.806	0.766	0.904

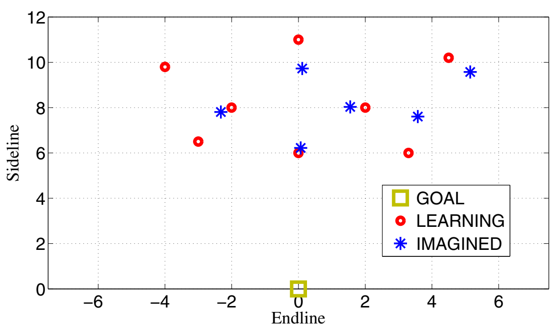

Fig. 4. Learn set: Eight Real and Imagined positions in cartesian coordinates

the environment with the 8 real and imagined positions respectively for training set and testing set. As the figures show, overall the robot is able to estimate its position in the environment to a good extent.

Table V reports the distance, evaluated using Cartesian coordinates, between the first and last positions in the imagined series and the real positions. As noticed in Table II the error is very high for position 8 of testing set and for position 2 and 7 of the learning set, positions where the robot fails to acquire the target because it makes a wrong move and the goal goes out of sight. In those cases, the first imagined position is quite good, but because of the wrong movement it is no longer able to see the goal, thus, it is not able to make the right following moves, and then to imagine the correct position in the field. Figure 6 and 7,

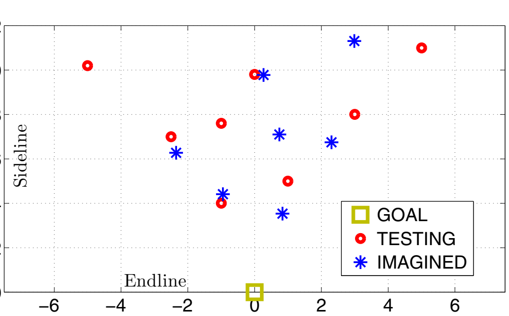

Fig. 5. Test set: Eight Real and Imagined positions in cartesian coordinates

TABLE V
DISTANCE OF IMAGINED POSITIONS WITH RESPECT THE REAL ONES

N	First imagined Learning	First imagined Test	Last imagined Learning	Last imagined Test
1	0.48	0.78	0.23	0.49
2	1.18	2.46	13.66	0.59
3	1.28	0.59	1.24	0.32
4	0.28	2.46	0.45	1.84
5	0.64	0.86	0.14	0.56
6	2.39	0.95	1.16	0.27
7	1.84	2.22	14.52	2.30
8	1.98	2.73	1.27	14.99
Avg	1.26	1.63	4.09	2.67
Avg without pos [2;7] / [8]			0.75	0.91

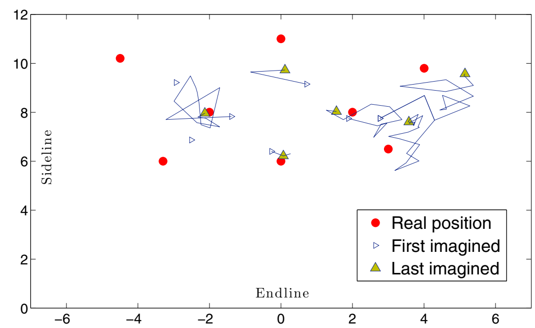

Fig. 6. Learn set: *imagined path* with first and last imagined positions compared with real positions in cartesian coordinates

where first and last imagined positions are shown along with the *imagined path* and the real position. The *imagined path* is the fictitious path composed by connecting all imagined positions according to the movements made. The *imagined paths* for positions with very high error are not depicted to avoid confusing the plot. Frm the image, it can be noted that the accuracy of imagined positions gradually improves while the robot performs the movement to center the goal and shoot. The improvement is of 0.51 meters for learning set and 1.76 meters for testing set. According to this result, the use of proprioceptive information, coming from the autonomous body movements, influences the robot imagination and, very often, helps it to better estimate its position in the field.

IV. CONCLUSION

In this work we presented the results of a preliminary study about mental imagery in cognitive robots. In particular we focused on spatial position estimation from proprioceptive and visual information, demonstrating that cognitive robotics can be an interesting and complementary framework to study mental imagery by means of real tasks. In this scenario, we believe that a better understanding of mental imagery in human will give the opportunity to apply such knowledge toward the development of artificial cognitive systems which will enable complex robots to better interact with the environment and refine their motor skill in an open-ended process. On the other end, recent humanoid robot platforms

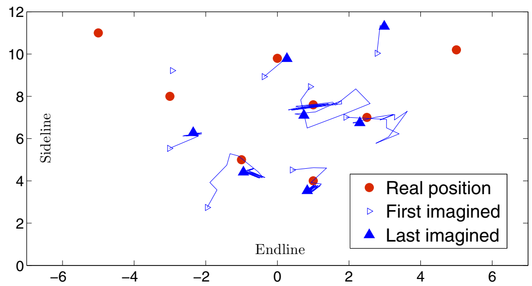

Fig. 7. Test set: *imagined path* with first and last imagined positions compared with real positions in cartesian coordinates

represent an important opportunity to build cognitive models of mental imagery and to regard the robotic systems as a test-bed for evaluating models of human imagery. Therefore, this work will be an additional step toward the incorporation of embodied cognition principles to the current research in Cognitive Science. In fact, the tight relation between mental simulations and real movements is a perfect example of embodied cognition, where the mind represents, or simulates, the body in action. The next step of this research will be to incorporate self-generated "mental" position in the training set and compare the performance with and without the help of the imagery training, aiming to replicate and to corroborate experimentally some of the data already known about the role of mental imagery in humans and sport practice, in particular. Moreover, we believe this work could be an additional step toward the incorporation of embodied cognition principles to the current research in Cognitive Science. The tight relation between mental simulations and real actions is a perfect example of embodied cognition, where the mind represents, or simulates, the body in action [13]. For this reason, the integration of traditional psychological studies with cognitive robotics models capable of embodied mental simulations can led to mutually fruitful insights and results, both for psychology and robotics.

References

[1] A. Clark and D. Chalmers, "The extended mind," *Analysis*, vol. 58, no. 1, pp. 10–23, 1998.

[2] P. R. and B. J., *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, November 2006.

[3] M. Wilson, "Six views of embodied cognition," *Psychonomic Bulletin & Review*, vol. 9, no. 4, pp. 625–636, 2002.

[4] A. Ishai, L. Ungerleider, and G. Haxby, "Distributed neural systems for the generation of visual images," *Neuron*, vol. 28, pp. 379–390, 2000.

[5] M. Jeannerod, "The representing brain. neural correlates of motor intention and imagery," *Behavioral & Brain Sciences*, vol. 17, pp. 187–245, 1994.

[6] J. Decety, J. M., and P. C., "The timing of mentally represented actions," *Behavioral & Brain Research*, vol. 34, pp. 35–42, 1989.

[7] G. Wei and J. Luo, "Sport expert's motor imagery: functional imaging of professional motor skills and simple motor skills," *Brain Res.*, vol. 1341, 2010.

[8] D. Smith, P. Holmes, L. Whitemore, D. Collins, and T. Devonport, "The effect of theoretically-based imagery scripts on field hockey performance," *Journal of Sport Behavior*, vol. 24, pp. 408–419, 2001.

[9] V. Tikhanoff, A. Cangelosi, P. Fitzpatrick, G. Metta, L. Natale, and F. Nori, "An open-source simulator for cognitive robotics research: the prototype of the icub humanoid robot simulator," in *PerMIS '08: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*. New York, NY, USA: ACM, 2008, pp. 57–61.

[10] M. J. L. Rumelhart, D. E., *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press, 1986.

[11] P. J. Werbos, "Consistency of hdp applied to a simple reinforcement learning problem," *Neural Networks*, vol. 3, no. 2, pp. 179–189, 1990.

[12] A. G. Di Nuovo, D. Marocco, S. Di Nuovo, and C. A., "Estimating spatial position from proprioceptive and visual information: A cognitive robotics model," in *Proceedings of tenth internation conference on epigenetic robotics*, Sweden, November 2010.

[13] T. Iachini, *Scienze della mente*. Bologna: Mulino, 2002, ch. Spazio, movimento e immagini mentali, pp. 176–184.