A Cognitive Robotic Model of Mental Rotation

Kristsana Seepanomwan[1], Daniele Caligiore[2], Gianluca Baldassarre[2], Angelo Cangelosi[1]
[1]Centre for Robotics and Neural Systems
Plymouth University, Plymouth, UK
{kristsana.seepanomwan, A.Cangelosi}@plymouth.ac.uk

[2]Laboratory of Computational Embodied Neuroscience, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR), Roma, Italy,
{daniele.caligiore, gianluca.baldassarre}@istc.cnr.it

***Abstract*—Mental rotation processes allow an agent to mentally rotate an image of an object in order to solve a given task, for example to make a decision on whether two objects presented with different rotational orientation are same or different. This article proposes a bio-constrained neural network model that accounts for the mental rotation processes based on neural mechanisms involving not only visual imagery but also affordance encoding, motor simulation, and the anticipation of the visual consequences of actions. The proposed model is in agreement with the theoretical and empirical research on mental rotation. The model is validated with a simulated humanoid robot (iCub) engaged in solving a typical mental rotation task. The results of the simulations show that the model is able to solve a mental rotation task and, in agreement with data from psychology experiments, they also show response times linearly dependent on the angular disparity between the objects. The model represents a novel account of the brain sensorimotor mechanisms that might underlie mental rotation.**

Keywords—Computational robotic model; neurorobotics; neural mechanisms; affordances and forward models; parietal/premotor/prefrontal cortex.

I. Introduction

In a typical mental rotation task, human participants are asked to make a decision on whether two objects presented with different rotational orientation are an identical or a mirror version of each other. The results show that the response times (RTs) linearly increase with the disparity angle between the orientations of the objects. The number of errors also increases as the disparity angle increases [1][2]. Since it was first described by Shepard and Metzler [1] in 1971, mental rotation has attracted enormous research interest in the field of cognitive psychology. This is in part due to the attempts to understand why object comparison using imagery seems to undergo to the same physical rules as overt rotation, considering that humans are capable of using imagery that is not limited by the laws of physics [3].

While early attempts to explain brain mechanisms underlying mental rotation relied upon visuo-spatial perception [1][4], recent behavioral and neuroscientific evidence suggest an important involvement of motor processes as well. In this respect, several behavioral works show interferences between action planning/execution and mental rotation processes [2][5][6]. Typically, during these experiments participants were asked to perform a classical mental rotation task [1] while performing a manual rotation on a custom joystick in both congruent and incongruent conditions with respect to the direction of rotation of the mental image. The results show that manual rotation affects mental rotation as RTs and error rate decrease only when the direction of the two rotations (manual and mental) is congruent. In addition, if the direction of manual rotations is opposite to the direction of mental rotation, mental rotation speed is slowed down [2][6].

Single cell recordings in the monkey's motor cortex also provide direct neural evidence for the involvement of motor processes in mental rotation [7]. In humans, several neuroscientific studies using different research techniques, such as transcranial magnetic stimulation (TMS), event-related potentials (ERPs), and functional magnetic resonance imaging (fMRI), show an involvement of lateral and medial premotor areas during mental rotation [8][9]. The fMRI study of Richter and colleagues [9], for example, shows a significant correlation of the hemodynamic response in lateral premotor areas with the RTs of participants involved in a typical mental rotation task [1]. This result suggests that mental rotation is an imagined (covert) object rotation rather than an image transformation relying exclusively upon visuo-spatial processing. This claim has been further confirmed by other studies (cf. [6][8][10]).

Importantly, despite these consistent results on the involvement of motor processes in mental rotation, a sound explanation of the specific brain mechanisms involving motor simulation that might underlie mental rotation processes is still lacking. One proposal that might help to understand the role of premotor areas during mental rotation pivots on the concept of affordance (i.e., the possible actions that objects and the environment offer to a certain agent) [11][12], and its brain correlates and models [13][14][15][16][17]. According to this framework, the visual presentation of objects triggers the activation, within the parietal-premotor circuits, of internal motor representations needed for the on-line guidance of actions over them [18][19]. In this respect, the activation of affordance representations might affect mental rotation processes as in brain it plays a key role in the first stage of motor preparation.

Another hypothesis on how premotor areas participate in mental rotation is based on the forward model theory of motor control [20][21][22]. According to this view mental rotation would involve the same motor areas and mechanisms used in the physical execution of active rotations of objects (e.g., manual rotations), and the imagined anticipation of their sensory consequences.

Both these hypotheses are interesting but do not explicitly consider that solving a mental rotation task is a complex skill requiring the coordinated action of several distinct cognitive processes. These processes include [8]: (a) stimulus encoding and mental image generation, (b) planning and execution of the mental rotation, (c) comparison (matching) of the rotated stimulus with the target stimulus, and finally (d) execution of the same/different response.

In this article we propose a computational model suggesting a neural operational hypothesis on how the information processing taking place in parietal and premotor areas might be involved in mental rotation. This operational hypothesis is based on the integration of affordances and forward model accounts for mental rotation discussed above. Combining these two perspectives within the model allows to deal with all the level of complexity required by a mental rotation task (not only the processes "a-b" indicated above (mental rotation proper), but also "c-d" (control and exploitation of mental rotation processes).

To this purpose, the model leverages on the computational model "TRoPICALS" [13][14][15] developed to study affordance compatibility effects [12]. TRoPICALS model is a good starting point to design a model on mental rotation as it reproduces some key functions of the parietal-premotor circuit, which are crucial for stimulus encoding and extraction of object affordances (process "a"). TRoPICALS also includes important features of the prefrontal-premotor circuit, pivotal for managing other aspects of mental rotation (processes "c" and "d"). However, it cannot perform mental image rotations as it lacks the needed feedback circuits. In this respect, to address the core mental rotation process (process "b") the model proposed here enhances the functions of TRoPICALS by developing two key new features. First, it is endowed with premotor-parietal feedback loops that allow it to implement mental rotation and sensory prediction based on forward models. Second, it is endowed with an improved visual and motor system allowing it to scale up to more realistic 3D environments and robotic setups.

The rest of the paper is organized as follows. Sec. II discusses the main features of the model, the learning algorithms used to train it, and the robotic set up used validate it. Sec. III presents and discusses the results. Sec. IV drives the conclusions and proposes future work to improve the model.

II. METHODS

A. Neural Architecture

The model proposed here represents an operational hypothesis on how visual and motor neural processes might interplay during mental rotation. To this purpose it extends some features of the TRoPICALS model [14]. Fig. 1 shows the model architecture which consists of three main parts corresponding to specific areas of the brain mainly involved during mental rotation tasks [8][9]: the parietal cortex (PC), the premotor cortex (PMC), and the prefrontal cortex (PFC). These areas are represented by distinct neural maps activated using population code methods [23][24]. The population code hypothesis postulates that information, e.g., on stimuli and actions, is encoded in brain on the basis of the activation of populations of neurons organized in neural maps having a broad response field. In particular, each neuron responds maximally to a certain value of the variables to encode, and then progressively less intensely to values (based on a Gaussian function).

The neurons of the PC map (32 x 32 neurons) encode the shape and orientation of the object that has to be mentally rotated [15]. The PMC consists of 2 neural maps PMC_1 (31 x 105 neurons) and PMC_2 (10 x 20 neurons), encoding motor programs related to different arm parts [20]. PMC_1 neurons encode a specific wrist posture of the robot corresponding to a specific object orientation encoded in PC. PMC_2 neurons encode the two different hand posture that the robot produces to accomplish the mental rotation results (i.e. to indicate if two objects are same or different). More in detail, the model works with 2 different types of object, each with 13 different orientations. Therefore, the neurons of the PMC_1 map encode 26 possible wrist postures.

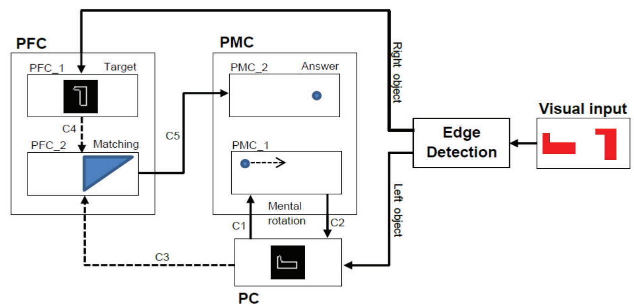


Figure 1. The model of mental image rotation. Each box represents the model's components. The arrows represent information flows from one component to another. The arrows accompanied by the letter "C" are the connections learned by SOM learning rule (dashed arrows) or by Hebbian learning rule (solid arrows).

The PFC has also 2 maps, implementing the working memory (PFC_1, 32 x 32 neurons) and the matching process area (PFC_2, 32 x 64 neurons) [25]. The visual input for the model is the image of a simulated camera of one of the eyes of a simulated iCub robot. This image goes through an edge detection module to extract edge information of the two objects shown in front of the robot. The edge information for the object on the left will be passed to the PC, while the one for the "target object" on the right will be for PFC_1. The target object is used as a reference for rotational purposes. The robot has to mentally rotate the object encoded by PC to check if it is the same or it is different with respect to the target object stored within PFC_1. PFC_2 is the core for the matching process. It is formed by a Kohonen self-organizing map (SOM) [26] which takes inputs from the PC and PFC_1. A major processing characteristic of the SOM is clustering. It transforms high dimensional inputs into low dimension as ones. Each input will then be represented in a unique area in the SOM map. We exploit this characteristic to create a neural map that represents pairs of stimuli. At the end of the matching process, PFC_2 neurons trigger PMC_2 activation whose neurons in turn encode the answering behaviour.

The mental rotation process is mainly based on the interactions between the PC and the PMC_1. Consistently with the concept of affordance, the visual features of the object (shape and orientation) encoded by the PC cause a specific cluster of neural activity in PMC_1. This pattern encodes the motor response to the seen object (i.e. a specific wrist rotation either clockwise or counter-clockwise, represented in terms of the posture assumed by the robot's wrist). Conversely, the PMC_1-PC circuit works as a forward model based on which a cluster of activity in PMC_1 causes a change of the image orientation in the PC.

B. Learning process

Connections between maps are trained using Hebbian learning and SOM competitive learning. Hebbian learning is widely accepted as a biologically plausible learning mechanism mainly involving cortical areas [27]. This learning mechanism underlies some developmental phenomena. One example is the critical period of learning [28], where synaptic efficacy cannot be modified and re-form after has been settled. The specific Hebbian learning method used in this model is the Oja rule [29], a Hebbian like equation that solves the problem of the basic Hebb rule causing a weights growing without bound.

At the beginning of the simulation, the weights of all the connections (C1, C2, C3, C4, and C5) are randomly set within the range [0, 0.1]. Then the simulated mental rotation experiment follows 4 steps:

- Stimulus encoding, assigning the edge information of the left stimulus to PC.
- Execution of the mental rotation, repeating the interaction between affordance (PC-PMC_1 circuit) and forward model (PMC_1-PC circuit; C1, C2 connections) processes.
- Comparison, performing the matching process of the mental image and the target image in the SOM map (PFC_2; C3, C4 connections).
- Answer triggering, executing of the same/different response (PFC_2-PMC_2 circuit; C5 connection).

The connections C1 are used to simulate affordance learning through the transformation of information from PC to PMC_1 [17]. The training set consists of pairs of the left stimulus (within PC) and a specific cluster of activity (Gaussian tuning curve) which represents the affordance provided in PMC_1 (i.e. the robot's wrist angle). For each pairs, the C1 connections are trained by using the following Hebbian learning rule:

$$\Delta w_{ij} = \eta\, a_i (a_j - w_{ij}) \tag{1}$$

$$w(t)_{ij} = w(t\text{-}1)_{ij} + \Delta w_{ij} \tag{2}$$

where Δw_{ij} denotes the weight's change from neuron i to neuron j, a_i and a_j denote activation potential of neuron i and j respectively, η denotes the learning rate which is set to 0.15, and $w(t)_{ij}$ is a weight value at a particular time step.

After training C1, an image of an object from the PC causes a specific cluster of activity in PMC_1, that represents a wrist posture. Through training the network learns how to rotate the robot's wrist corresponding to the orientation of a seen object.

The connection C2 is responsible for forward model learning. In contrast to the affordance processing, this connection causes the formation of an image representation in PC from the cluster of activity in PMC_1. For instance, a cluster of activity in PMC_1 that is caused by an image of object rotated 90 degrees in PC (during affordance processing) causes a 75 degrees rotated image back in PC. This training strategy allows the network to create a series of rotating images. Note that the training set causes an image to gradually change to become the same as an image of object of 0 degrees. This corresponds to the central position in PMC_1 map, which refers to the target position. When a rotation of input image is greater than 0 degrees, the image will be rotated on the right (clockwise). In contrast, if the angle is less than 0 degrees, the image will be rotated on the left (counter-clockwise). The C2 connections are also trained with the Hebbian learning rule used to train C1 (Eq. 1).

The process of mental image rotation consists in the repetition of the interaction between the affordance process (connection C1) and the forward model process (connection C2), until an image in the PC reaches the 0 degrees target rotation. Each cycle of the interaction causes a rotated image, which can be considered a mental image because the actual input object orientation does not change.

The connections from PC and PFC_1 to SOM PFC_2 (C3, C4) are responsible for the matching process. When the network generates a mental image in the PC, having a 0 degrees rotation, then the process of learning is triggered. The connections link two maps, one is PFC_1 (target image), which is set at the beginning of the simulation, and another is PC (the mental image). A training set for PFC_2 is a combination of all the possible neural representations for the stimuli of each input. A neural activity in PFC_2 forms a salient cluster with respect to the two specific inputs. As there are two possible images in each map, four clusters will be formed. To train PFC_2, the following SOM learning rule was used:

$$w(t)_i = w(t\text{-}1)_i + \Theta(t\text{-}1)_i\, \eta(t\text{-}1)_i\, (v(t\text{-}1)_i - w(t\text{-}1)_i) \tag{3}$$

where $w(t)_i$ denotes current weight value of neuron i at time t, $w(t\text{-}1)_i$ denotes an old weight value of the neuron i, Θ denotes the amount of influence on distance between neuron i and the best matching neuron in a map, η denotes the learning rate which is set to 0.05, v denotes input value to the neuron i. Note that, Θ and η decrease over time.

The PFC_2 SOM map is trained in advanced. In this way, a response of PMC_2 can be fixed for each input couple from PC and PFC_1.

The answer triggering process uses the connection C5 from PFC_2 to PMC_2. When two images are "similar" the robot chooses the "YES" answer, otherwise it chooses the "NO" answer. The term “similar” means “it is approximately the same”. The mental rotation ends when the position of cluster of activity in PMC_1 is close to the central position. The most salient cluster in PFC_2 is used to produce the answer. Given the four possible combinations of inputs in the matching process, two of them are responsible for a "SAME" answer, while the remaining two for the "MIRROR" answer. Therefore, two regions in PFC_2 with respect to the same image from the PC and PFC_1 cause one cluster in PMC_2. While two other regions within PFC_2 represent different images of the two input maps. In this process, PMC_2, is responsible for the answer triggering, the motor response to press two answer buttons or to produce some utterance such as “YES” or “NO”. In the current version of the model this motor command is still not used to supply a control signal for the iCub but is directly interpreted as the response of the system.

After learning, an action potential of each neuron in the PMC_2 map is calculated by using a dynamic competition method [27]. As the connections within a neural map are based on an all-to-all pattern, each neuron in the map sends/receives signals to/from every neuron. The dynamic competition process causes dynamic activities within the map, based on a distance between neurons following the rule of long-range inhibition and short-range excitation. Neighbouring neurons which are activates with high potential will receive excitatory signals and tend to form clusters of activity. In contrast, the neurons which are far from the active neuron in the neural space will receive an inhibition signal and their action potential will be depressed.

The dynamic competition is also used as a method to calculate an agent’s response time, e.g. to compare the model results with reaction time data in psychology experiments. Unlike a simple feed-forward process in layered neural networks, the dynamic competition process will be repeated until the action potential of at least one neuron in the neural map reaches a specific threshold. This process can be used to calculate the response time based on the action potential of an individual neuron that is most sensitive to a particular input. In detail, the number of repeating dynamic competition processes was recorded and used as simulated response time. One cycle of repeating the process will be assumed to be equal to 1 millisecond [13].

C. *The simulated participant (the iCub robot)*

According to the view of embodied cognition [31][32], our cognitive capabilities to recognize and understand things have been shaped by the interaction processes between body, brain and environment. In addition, cognition is based on internal representations and simulations of real world actions and our perception [33].

Cognitive robotics platforms, such as humanoid robots, are being increasingly used to model embodied cognition and cognitive development in humans by means of embodiment [13][14][34]. Following this approach, a simulation model of the humanoid robot iCub was used to model psychological experiments on the embodiment bases of mental rotation.

The iCub robot [35] is an open source robotic platform built for studying cognitive development in humans. It looks like a human 3-5 years old child, and has an excellent layout of humans’ body structures and movements. The iCub has been widely used as cognitive tool in many robotics laboratories. An iCub’s simulator [36] (Fig. 2), which replicates the same body and control scheme of the real iCub, will be used in this work. It provides visual perception via simulated cameras and can perform actions corresponding to specific motor commands.

Each arm of the iCub has 16 joints. This work uses the joint number 5 of the right arm which directly affects the robot wrist’s angle. If the robot holds an object with the right hand, rotating the wrist will change only orientation in the object plane.

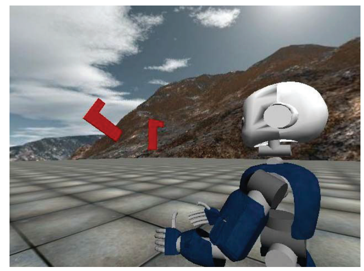


Figure 2. The iCub simulator and its environment

D. *Stimuli and Simulated Mental Rotation Task*

The visual stimuli use an abstract object, colored in red, similar to an upside down letter L as shown in Fig. 3. In this experimental set up, two versions of these stimuli are used, each producing a mirror image of the other, and will be called object-A and object-B. The objects are displayed in the space in front of iCub (Fig. 2). During the process of affordance training, only one stimulus is shown in the left position, with the experimenter varying the orientation of the object and assigning a corresponding target position of the robot's wrist angle. In the testing session, two stimuli are displayed in the left and in the right positions. In each trial, the rotation of the left image is systematically varied, while the right one is presented with a 0 degrees orientation and can involve the two objects A and B.

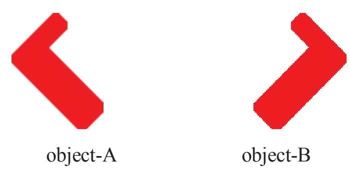


Figure 3. The two stimuli used for the simulated mental rotation task. Both stimuli are colored in red for edge detection.

The edge detection method is used as an early visual processing stage. The image is centred on a single object, and the red color filter is applied. The edges of the object are extracted with the Canny edge detection technique [37], using the OpenCV library. The output from the edge detection process consists of binary data which can be directly assigned as an activity level to PC and PFC_1 at the beginning of the simulation. Note that the eyes’ position of the iCub was fixed, with the object centred on the fovea throughout the experiment.

Regarding the motor response, there is a limitation of the iCub's wrist angle, which can rotate in the range of [-90; 90] degrees. Counter-clockwise orientations are indicated by positive values, while clockwise orientations are indicated by negative values. For example, in Fig. 3 object-A has a 45 degrees orientation while object-B a -45 degrees one.

III. RESULTS

The right object is always shown at a 0 degree rotation, while the left object can vary in orientation between 90 and -90 degrees. Therefore the maximum angular disparity between the two stimuli is 90 degrees. Varying them by 15 degrees (0, 15, 30, 45, 60, 75), as we did, this will typically require a maximum mental rotation in the map PMC_1 of 6 steps. However, in the experiment the maximum number of rotation cycles is set to 10 as in some cases the model cannot rotate the image to a preferred orientation at the first cycle, thus requiring extra rotations. When the number of rotation cycles is equal to 10, it indicates that the model cannot correctly perform the mental image rotation of the left stimulus and will be forced to do the next step (matching process) by using the last image. The interaction between affordance and forward model processes leads the model to obtain a linear relationship between the angular disparity and a number of steps used in rotation.

The experiment is conducted using two groups of inputs, one for a recognition test and another for generalization test. In the recognition test, orientations of the left stimulus are the same as in the training set by varying 15 degrees per pattern from 90 to -90. As there are two possible objects and each of them can have 13 possible orientations, this test has exactly 13x4=52 different pairs of stimuli to be used as input. The generalization test refers to testing the model with unseen orientations. The left stimulus in this test changes 5 degrees from 90 to -90 but skip the cases of repeated values of the previous test. Therefore, the generalization test has (37x4)-52 = 96 pairs of stimuli to be used as input. Both tests were repeated 52 times to record the consistency of the model performance. The result showed in Fig. 5a is a series of mean values of response time of the recognition test.

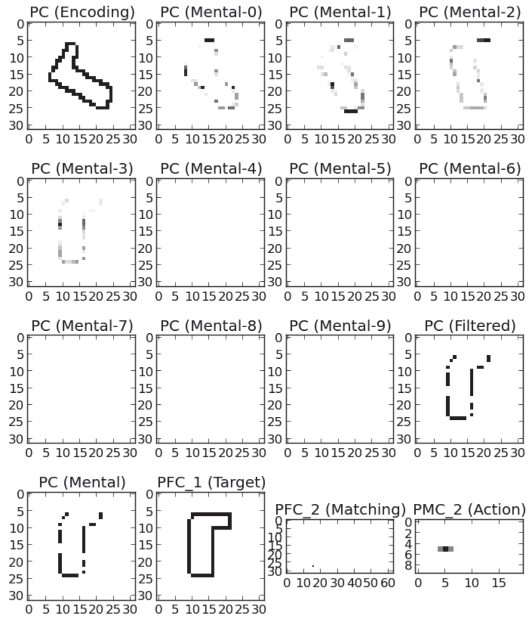
Fig. 4a shows the mental rotation steps (PC) and the matching (PFC_2) and answering (PMC_2) processes for a successful trial. In this example the mental rotation process takes 4 steps to rotate an image of a stimulus in 60 degrees to an image of stimulus in 0 degrees. The mental rotation process ends when the rotated image reaches 0 degrees orientation. After that, the matching process within PFC_2 is performed by using as input the neural activity of target image in PFC_1, and the rotated image in PC. The neural activation representing the matching process within PFC_2 is showed in the third column of the last row in Fig. 4a. The answering process of PMC_2, is indicated in the fourth column of the last row on Fig. 4a. The cluster of activity formed in the left side of the map will cause the answer "YES" to be chosen. The blank panels indicated that the rotational steps needed in this sample are less than 10.

In contrast, Fig. 4b shows one case in which the model cannot rotate the left stimulus of -90 degrees of object-A into the 0 degrees default position. The model fails to rotate the image within 10 cycles, and has to do the matching process by using the last (un-rotated) image in PC. This scheme is similar to a guessing process in human subjects when the time to do mental rotation task is over. When the model fails to rotate the image after 10 cycles: each cycle, the image in the PC is the same. This case might be caused by a similarity effect of the edge information of objects in the training set. Indeed, the edge information of object-A and object-B in 90 and -90 degrees are similar in pattern as they mostly lay on the horizontal axis in the centre of the map. This means the model has to learn to match 4 similar inputs related to 4 separated clusters (on the left-most or right-most of the map). Hence, the model cannot learn to match the case of object-A in -90 degrees to a correct cluster. Therefore, when the left stimulus is object-A in -90 degrees, the affordance process causes a cluster of activity in PMC_1 which is not exactly the preferred position (it is a nearby position). Then the forward model, using that cluster, causes the same image back in the PC.

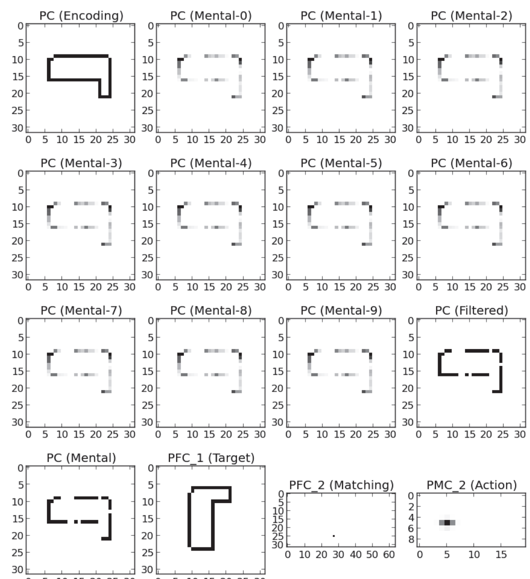
The model has successfully reproduced the findings with human subjects in a typical mental rotation task. Indeed, in experiments with human RTs profile often shows an inverse v-shaped profile as the one found in the simulations we ran. Fig. 5a shows the RTs profile of the recognition test where the model performed simulated mental rotation tasks with different orientations of the left stimulus. From the figure it is possible to observe how RTs increase as the angular disparity increases.

As indicated by the RTs profile, the mental rotation performance with the object-A produces higher RTs than with object-B. This characteristic is affected by the training strategy that feed a sequence of patterns in the training sets. The patterns of object-A are fed to the network in a training period before object-B. This makes the model more sensitive to object-B than object-A. However, this characteristic might be prevented by using a random feed, instead of sequential feed of training patterns. This behaviour of the model might be also considered a prediction for a possible experiment with real subjects where these are allowed to manipulate different objects in different times before the test.

The recognition test achieves 98% correct answers (51 out of 52) on rotating object A and 100% on rotating objects B. While in the generalization test with unseen orientations, the success in rotation is 95.8% (92 out of 96) and 96.8% (93 out of 96) respectively for object B and object A. These results are always the same over 52 trials. The overall successful rate is 96.6%. The angular disparity between the two stimuli affects the response time of the model in the same fashion as human subjects. Increasing the difference in degrees of rotation causes the model to require a higher number of rotating cycles: this increase the simulated response time. The experimental result show that the model can rotate most of the possible pairs of objects except for one case of object-A in -90 during recognition test, 2 cases of object-A in -85 and 70, and 2 cases of object-B in 55 and 70 degrees during learning stage. Dash circles in both graph of Fig. 5 are used to point an orientation of the stimulus that cannot be rotated. When an image of that stimulus is shown, the model cannot change it into a new preferred image rotated 15 degrees.



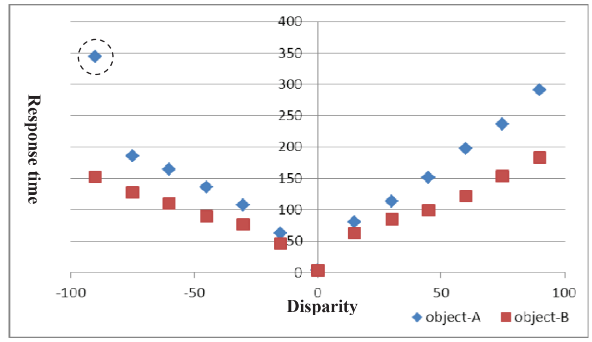
a)



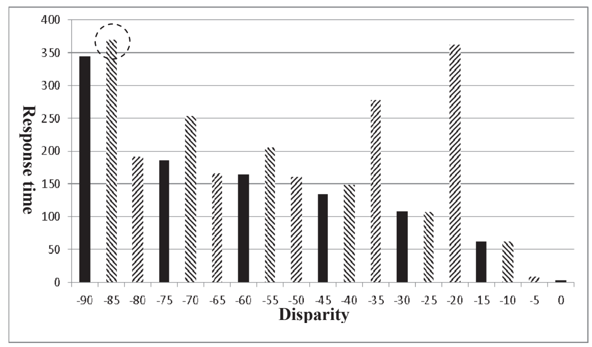
b)

Figure 4. Mental image rotation steps. a) Rotational steps in the case that the model is able to create a series of image changes to reach the 0 degrees default orientation; b) the model is unable to rotate the seen object. The matching and answering processes are represented by the neural activation of the two bottom right side maps respectively.

The Fig. 5b reports the comparison of RTs profile between recognition test and generalization test on object-A. The figure indicates that some unseen orientations in the learning test took more RTs than one in the recognition test even the disparity is smaller. At each step of image rotation, a new image is gradually changed approaching the 0 degrees target object rotation. When the model is shown an unseen object orientation, this causes longer reaction times due to the intermediate angle of rotation not matching the discrete 15 degrees increment images shown during training. For example, when the model is shown an image with 55 degrees rotation, the RTs will be greater than performing of 60 degrees one. This is because first the model keeps rotating the image of 55 degrees stimulus into an image of 60 degrees stimulus for some steps (55 is closer to 60 than 45 according to training patterns), and only after this it goes back to follow the incremental 15 rotating strategy.



a)



b)

Figure 5. RTs profiles of the simulated mental rotation task. a) The rotation of different stimuli affects the RTs profile; b) Series of some unseen (gray bars) and seen (black bars) stimulus orientations of object-A; notice that some unseen orientations (-85, -70, -55, -35. -20) show greater RTs than ones in the training set (-90, -75, -60, …).

The reasons leading the model to produce a noisy RTs profile for the generalization test (Fig. 5b) could be explained as follows. If in the first cycle the starting image in PC is unseen the feed-forward process, through the connection C1, could activate a neural cluster within PMC_1 representing an unpredictable position. As a result, the mismatched salient cluster in the map PMC_1 creates an incorrect image back to the PC. These processes are repeated, by chance, and the mental rotation ends when the position of the cluster of activity in PMC_1 is close to the central position. This process leads the model to produce a noisy RTs profile for the generalization test as shown in Fig. 5b. To sum up, the similarity of unseen object orientations to the training patterns is the main explanation of this effect. However, this is a common effect that can be found when working with neural networks. One way to solve this effect is to use more precise training patterns.

IV. CONCLUSION AND FUTURE WORKS

The model proposed in this article accounts for the mental rotation processes based on neural mechanisms involving visual imagery, affordance encoding and forward models processing. In this respect, the proposed approach is in agreement with the theoretical and empirical research on mental rotation about the interplay between covert motor processes and the creation of mental image during mental rotation task. Remarkably, the model is validated within the simulated humanoid robot iCub engaged in solving a typical mental rotation task.

The model also presents some limits which are, however, all addressable in future work. Here we briefly discuss the main limits and propose a solution for each of them. First of all, in the current version of the model mental rotation mainly depends on the interplay between affordance and forward model processes, ignoring the role of the wrist proprioceptive signals. This means that the robot wrist movements do not influence the mental rotation processes. Recent research [38][39] points out that the performance of mental rotation tasks can be improved by the assistance of hand movements, or gestures called "co-thought gestures". These studies also suggest that spontaneous gestures during performing mental rotation task provide a rich sensorimotor experience to the solving strategy in human subjects. Gestures improve the internal representation of a spatial transformation of objects. Following this hypothesis, we will modify the current model by adding proprioceptive units that should act as an internal representation of gestures or hand movements. More in details, we will introduce a more direct effect of proprioception on mental rotation by modifying the parietal area by introducing a somatosensory map (SS) whose neurons encode the proprioceptive signal. SS might be a dynamical field map [30] combining the forward model signal with the proprioceptive signal. When the two signals are different there would be an interference effect and the dynamical competition would take more time to be solved (increasing RTs). The new version of the model should be able to account for other data which link overt movements and mental rotations [5][6] (cf. Introduction).

Secondly, humans can create mental images and perform mental image rotation on a variety of objects, even on unseen abstract object [1][2]. Unlike humans, the model proposed here can work only with the objects of a training set. While it should be able to work with any kind of object. This will be done by separating the object orientation from the "object identity". We think that this could be possible using an inferotemporal cortex (IT) map whose neurons encode objects independently of their orientation, similarly to what happens in humans. This map could be connected to PC together with SS (encoding the posture proprioceptive signal): this would allow PC to encode the combination of object identity (IT) and particular wrist posture (SS) so as to allow the system to imagine the rotation of any type of object after being suitably trained.

Overall, also considering these possible improvements, the proposed neuro-robotic model of mental rotation provides a useful computational framework to study the integration between mental rotation capabilities and embodied cognition.

ACKNOWLEDGMENT

This research received funds from EU funded projects Robot-Era (FP7-IP-288899), Poeticon++ (FP7-STREP-288382), a grant from the Thai government, and "IM-CLeVeR–Intrinsically Motivated Cumulative Learning Versatile Robots", contract no. FP7-IP -231722.

REFERENCES

[1] R. Shepard and J. Metzler, "Mental rotation of three dimensional objects," Science 1971. 171(972):701-3.

[2] M. Wexler, S.M. Kosslyn, and A. Berthoz, "Motor processes in mental rotation," Cognition, 68 (1), 1998, pp. 77-94.

[3] S.M. Kosslyn. Image and the Brain. The Resolution of the Imagery Debate. MIT Press/A Bradford Book, Cambridge 1994.

[4] M.C. Corballis and R. McLaren, "Interaction between perceived and imagined rotation," J. Exp. Psychol. Hum. Percept. Perform., 8, 1982, pp. 215-224.

[5] A. Wohlschläger and A. Wohlschläger, "Mental and manual rotation," J. Exp. Psychol. Hum. Percept. Perform., 242, 1998, pp. 397-412.

[6] A. Wohlschläger, "Mental object rotation and the planning of hand movements," Percept. Psychophys., 63, 2001, pp. 709-718.

[7] A.P. Georgopoulos, J.T Lurito, M. Petrides, and A.B. Schwartz, J.T. Massey, "Mental rotation of the neuronal population vector," Science., 243, 1989, pp. 234-236.

[8] C. Lamm, C. Windischberger, E. Moser, and H. Bauer, "The functional role of the dorso-lateral premotor cortex during mental rotation An event-related fMRI study separating cognitive processing steps using a novel task paradigm," NeuroImage., 36, 2007, pp. 1374-1386.

[9] W. Richter, R. Somorjai, R. Summers, M. Jarmasz, R.S. Menon, J.S. Gati, et al. "Motor area activity during mental rotation studied by time-resolved single-trial fMRI," J. Cog. Neurosci., 12, 2000, pp. 310-320.

[10] C. Lamm, P.H.S. Fischmeister, and H. Bauer, "Individual differences in brain activity during visuo-spatial processing assessed by slow cortical potentials and LORETA," Cogn. Brain. Res., 25, 2005, pp. 900-912.

[11] J.J. Gibson, "The Ecological Approach to Visual Perception," Boston: Houghton-Mifflin, 1979.

[12] M. Tucker and R. Ellis, "The potentiation of grasp types during visual object categorization," Vis. Cogn., 8, 2001, pp. 769-800.

[13] D. Caligiore, A.M. Borghi, D. Parisi, G. Baldassarre, "Affordances and compatibility effects: A neural-Network computational model," In J. Mayor, N. Ruh, K. Plunkett (Eds.), Connectionist Models of Behaviour and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop , Singapore: World Scientific, 2008, pp. 15–26.

[14] D. Caligiore, A.M. Borghi, D. Parisi, G. Baldassarre, "TRoPICALS: A computational embodied neuroscience model of compatibility effects," Psych. Rev., 117, 2010, pp. 1188-1228.

[15] D. Caligiore, A.M. Borghi, D. Parisi, R. Ellis, A. Cangelosi, G. Baldassarre, "How affordances associated with a distractor object affect compatibility effects: A Study with the computational model TRoPICALS," Psych. Res., 2012. DOI: 10.1007/s00426-012-0424-1

[16] G. Rizzolatti and L. Craighero, "The mirror-neuron system," Annu. Rev. Neurosci., 27, 2004, pp. 169-192.

[17] A. H. Fagg and M. A. Arbib, "Modeling parietal-premotor interactions in primate control of grasping," Neural Networks, 11 (7-8), 1998, pp. 1277-1303.

[18] S.T. Grafton, L. Fadiga, M. A. Arbib, and G. Rizzolatti, "Premotor cortex activation during observation and naming of familiar tools," NeuroImage, 6, 1997, pp. 231-236.

[19] J. Grèzes, and J. Decety, "Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis," Hum. Brain. Mapp., 121, 2001, pp. 1-19.

[20] D.M. Wolpert, and M. Kawato, "Multiple paired forward and inverse models for motor control," Neural Networks, 11, 1998, pp. 1317-1329.

[21] R. Grush, “The emulation theory of representation: motor control, imagery, and perception,” Behav. Brain Sci., 27 (3), 2004, pp. 377-396.

[22] R. C. Miall, “Connecting mirror neurons and forward models,” Neuroreport, 14 (17), 2003, pp. 2135-2137.

[23] A. Pouget, P. Dayan, and R. S. Zemel, “Inference and computation with population codes,” Annu. Rev. Neurosci., 26, 2003, pp. 381-410.

[24] S. Deneve, P. Latham, and A. Pouget, “Reading population codes: A neural implementation of ideal observers,” Nat. Neurosci., 2 (8), 1999, pp. 740-745.

[25] J.M. Fuster, "The prefrontal cortex--an update:time is of the essence," Neuron, 30 (2), 2001, pp. 319-333.

[26] T. Kohonen, Self-organizing maps, 2nd ed. Berlin, Germany: Springer-Verlag, 1997.

[27] K. Doya, “Complementary roles of basal ganglia and cerebellum in learning and motor control,” Curr. Opin. Neurobio., 10, 2000, pp. 732-739.

[28] Y. Munakata and J. Pfaffly, “Hebbian learning and development,” Dev. Sci., 7, 2004, pp. 141-148.

[29] E. Oja, “A simplified neuron model as a principal component analyzer,” J. Math. Biol., 15, 1982, pp. 267-173.

[30] W. Erlhagen and G. Schöner, “Dynamic field theory of movement preparation,” Psychol. Rev., 109 (3), 2002, pp. 545-572.

[31] D. Pecher, and R.A. Zwaan, “Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking,” Cambridge University Press, 2005.

[32] G. Pezzulo, L.W. Barsalou, A. Cangelosi, M.H. Fischer, K. McRae, and M. J. Spivey. “The mechanics of embodiment: a dialog on embodiment and computational modelling,” Front. Psychol., 2(5), 2011, pp.1-21.

[33] L. Barsalou, “Perceptual symbol systems,” Behav. Brain Sci. 22, 1999, pp. 577-609.

[34] A. Cangelosi and M. Schlesinger, Developmental robotics, MIT press, in press.

[35] G. Sandini, G. Metta, and D. Vernon, “The iCub cognitive humanoid robot: An open-system research platform for enactive cognition,” in 50 Years of AI, M. Lungarella, Ed. et al. Berlin, Germany: Springer-Verlag, 2007, pp. 359–370.

[36] 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,” Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, 2008, pp. 57-61.

[37] F. J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., 8 (6), 1986, pp. 679-698.

[38] M. Chu and S. Kita, “Spontaneous gestures during mental rotation tasks: insights into the microdevelopment of the motor strategy,” J. Exp. Psychol. Gen., 137, 2008, pp. 706–723.

[39] M. Chu and S. Kita, “The nature of gestures’ beneficial role in spatial problem solving,” J. Exp. Psychol. Gen., 140, 2011, pp. 102–116.