Mental practice and verbal instructions execution: a cognitive robotics study

Alessandro G. Di Nuovo  
Davide Marocco, Angelo Cangelosi,  
Plymouth, University  
Drake Circus, Plymouth PL4 8AA, UK

Vivian M. De La Cruz,  
Università degli Studi di Messina,  
via Concezione n. 6, 98122, Messina, Italy

Santo Di Nuovo,  
Università degli Studi di Catania,  
Via Biblioteca 4, 95124, Catania, Italy

Abstract—Understanding the tight relationship that exists between mental imagery and motor activities (i.e. how images in the mind can influence movements and motor skills) has become a topic of interest and is of particular importance in domains in which improving those skills is crucial for obtaining better performance, such as in sports and rehabilitation. In this paper, using an embodied cognition approach and a cognitive robotics platform, we introduce initial results of an ongoing study that explores the impact linguistic stimuli could have in processes of mental imagery practice and subsequent motor execution and performance. Results are presented to show that the robot used, is able to “imagine” or “mentally” recall and accurately execute movements learned in previous training phases, strictly on the basis of the verbal commands issued. Further tests show that data obtained with “imagination” could be used to simulate “mental training” processes such as those that have been employed with human subjects in sports training, in order to enhance precision in the performance of new tasks through the association of different verbal commands.

Keywords- motor imagery, embodied cognition, cognitive robotics, mental training, recurrent neural network.

I. INTRODUCTION

The processes behind the human ability to create mental images of events and experiences have recently become an object of renewed interest in cognitive science (e.g. [1]). Brain-imaging studies have shown that the same brain areas that are activated when seeing are also activated when recalling images [2]. It has also been shown that primary motor cortex is activated during the production of motor images as well as during the production of active movement [3,4]. Evidence also exists that suggests that language comprehension processes, also involve the activation of specific motor regions of the brain depending on the linguistic constructs heard by subjects. During the reading of verbs related to concrete action, for example, it has been found that the recruitment of the effector-specific regions in primary motor or pre-motor cortex is similar to the activation found in those areas when moving the effector that is most involved in those actions [5]. The first type of evidence, would be partly in line with what has been the controversial yet historically dominant interpretation given by philosophers and psychologists alike to the term “mental images”, that is, that they are a type of “inner pictures” in the mind. The second type, suggests that despite this general tendency to attribute mental images to being quasi-visual phenomena, it would be more appropriate to consider them as quasi-perceptual experience in any sensory mode or combination of sensory modes that is experienced in the absence of the actual stimuli. We consider mental imagery to be a multimodal mental simulation that activates the same, or very similar sensorial modalities, that are activated when we interact with the environment in real time. This could explain why both brain hemispheres and several functional areas (i.e., perceptual, linguistic, and motor) are involved in parallel when a mental image is produced and/or processed [6, 7].

A number of recent studies have shown that the capacity to mentally simulate actions, as well as their end results, plays an extremely important part in our ability to plan and execute actions as well as understand those of others. Mental simulation is often mediated by language. Language is instrumental in the vicarious experiencing of the events characteristic of mental simulation. It describes real or imagined events, guides our attention and orchestrates the retrieval of experiential traces of people, places, objects, actions, and events [8]. Even at the level of the single word, linguistic constructs have been linked to the sensorimotor memory traces that form the basis of mental simulation.

In sports, beneficial effects of mental training for performance enhancement in athletes are well established and a number of studies in the literature have explored and reviewed them as well as presented new training principles, e.g [9, 10]. For example in [11], a cognitive behavioral training program was implemented to improve the free-throw performance of college basketball players, finding improvements of over 50%. Furthermore, the trial in [12], where mental imagery was used to enhance the training phase of hockey athletes to score a goal, have shown that imagery practice helped in obtaining better performance. Despite these results, and the ample evidence that suggests that mental imagery, and in particular motor imagery, contributes to the enhancement of motor performance, the topic is still a relatively new subject of research, for example ([13]) investigated the effect of mental practice to improve game plans or strategies of play in a trial with 10 female basketball players. Results of the trial support the assumption that motor imagery may lead to better motor performance in open skills when compared to the no-practice condition.
In this work, we deal with motor imagery and how verbal instruction may evoke the ability to imagine movements, already seen before or new ones obtained by combination of past experiences. These imagined movements either replicate the expected new movement required by verbal commands or correspond in accuracy to those learned and executed during training phases. Motor imagery is defined as a dynamic state during which representations of a given motor act are internally rehearsed in working memory without any overt motor output [14].

Our approach addresses embodied cognition using humanoid robots. The concept of embodied cognition affirms that the nature of intelligence is largely determined by the body and by how it interacts with the world. The use of humanoid robots in studies of embodied cognition present many new opportunities in studying how cognition develops in humans. They allow us to identify and simulate aspects of cognition incrementally, in artificial systems that can learn through the interactions their bodies have with their environment. 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 actuator arrangements determine a highly redundant morphological structure, which is difficult to control.

Understanding the tight relationship that exists between mental imagery and motor activities (i.e. how images in the mind can influence movements and motor skills) is also of particular importance in domains in which improving those skills is crucial for obtaining better performance, such as in sports and rehabilitation. In [15] the authors, as part of an ongoing research program on mental imagery using cognitive robotics, explored how the relationship between spatial mental imagery practice in a training phase, could increase accuracy in sports related performance. The focus was on the capacity of the robot to estimate after a period of training with proprioceptive and visual stimuli. From a technological point of view, this research aims to reach a better understanding of mental imagery in humans, in order to derive engineering principles for the development of artificial cognitive systems capable of interacting better with their environment and of refining their motor skills in an open-ended process.

In this paper we extend the work mentioned above, by focusing on the integration of auditory stimuli in the form of verbal instructions, to the motor stimuli already experienced by the robot in past simulations. Simple verbal instructions are added to the training phase of the robot, in order to explore the impact that linguistic stimuli could have in its processes of mental imagery practice and subsequent motor execution and performance. In particular, we tested the ability of our model to use imagery to execute new orders, obtained combining two single instructions. This effort has been inspired by embodied language approaches, that are based on evidence that language comprehension is grounded in the same neural systems that are used to perceive, plan, and take action in the external world.

II. MATERIALS AND METHODS

The robotic model used for the experiments is the iCub humanoid robot controlled by a recurrent artificial neural network. The iCub is an open-source humanoid robot platform designed to facilitate developmental robotics research (e.g. [16]). This platform is a child-like humanoid robot 1,05m tall, with 53 degrees of freedom distributed in the head, arms, hands and legs. A computer simulation model of the iCub has also been developed. The simulated iCub has been designed to reproduce, as accurately as possible, the physics and the dynamics of the physical iCub. The simulator allows creating realistic physical scenarios in which the robot can interact with a virtual environment. Physical constraints 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 neural system that controls the robot is a three layer Recurrent Neural Network (RNN) with the architecture proposed by Elman [17]. The Elman RNN adds in the input layer a set of "context units", directly connected with the middle (hidden) layer with a weight of one (i.e., directly copied). At each time step, the input is propagated in a standard feed-forward fashion, and then a learning rule is applied. The fixed back connections result in the context units always maintaining a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). This creates an internal state of the network, which allows it to exhibit dynamic temporal behavior. To model mental imagery the outputs related with the motor activities are redirected to corresponding inputs.

The main difference between a standard Feed-Forward Neural Network and the RNN is that, in the latter case, the training set consists in a series of input-output sequences. The RNN architecture allows the robot to learn dynamical sequences of actions as they develop in time. The goal of the learning process is to find optimal values of synaptic weights that minimize the error, defined as the error between the teaching sequences and the output sequences produced by the network. Specific neurons, one for each verbal instruction, were included in the input layer of the RNN in order for it to take into account these commands, while the sensorimotor information is directed to the rest of the neurons in the input layer. The RNN architecture implemented, as presented in Figure 1, has 4 output units, 20 units in the hidden layer, and 27 units in the input layer, 7 of them encode the proprioceptive inputs from the robot's joints and 20 are the context units, i.e. are back links of hidden units, they only copy the value from output of upper unit to the input of lower unit. As learning process we used the classic back-propagation algorithm [18], the goal of which 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:

\[ E = \frac{1}{2} \sum_{i=1}^{p} [y_i - t_i]^2 \]  

(1)

where \( p \) is the number of outputs, \( t_i \) is the desired activation value of the output unit \( i \) and \( y_i \) is the actual activation of the same unit produced by the neural network, calculated using equation (2) and (3). During the training phase, synaptic weights at learning step \( n+1 \) are updated using the error calculated at the previous learning step \( n \), that in turn depend on the error \( E \). Activations of hidden and output units \( y_i \) are calculated by passing the net input \( u_i \) to the function, as it is described in equations (2) and (3):

\[ u_i = \sum j w_{ij} \cdot k_j \]  

(2)

\[ y_i = \frac{1}{1+e^{-u_i}} \]  

(3)

where \( w_{ij} \) is the synaptic weight that connects unit \( i \) with unit \( j \) and \( k_i \) is the bias of unit \( i \).

The back-propagation algorithm to update link weights and neuron biases, with a learning rate (\( \alpha \)) of 0.2 and a momentum factor (\( \eta \)) of 0.6, according to the following equation (4):

\[ \Delta w_{ij}(n+1) = \eta \Delta y_j + \alpha \Delta w_{ij}(n) \]  

(4)

where \( \Delta y_j \) is the error of the unit \( j \) calculated at the previous learning step \( n \). To calculate the error at first step of the back-propagation algorithm, initial values of back links are initialized to one. The network parameters are initialized with randomly chosen values in the range \([-0.1,0.1]\). In our experiments the learning phase was unstable, for this reason it was run only for 10,000 epochs, but we saved weights and biases of the epoch with the best mean square error.

After the learning phase, we tested the ability of the RNN architecture to model “mental” imagery, adding other back connections from motor outputs to motor inputs, at the same time connections from/to joint encoders and motor controllers are deactivated. This setup is presented in Figure 1, where red connections are active in “imagery mode” only, while green connections are deactivated in “imagery mode”.

III. EXPERIMENTS

Using the material and methods described in section 2, with the iCub simulator we performed two experiments:

1. The first experiment aimed to evaluate the ability of the RNN to model artificial mental imagery. It was divided into two phases: in the first phase the network was trained to predict its own subsequent sensorimotor state. The task was to throw in different directions (forward, left, right, back) a small object that was placed in the right hand of the robot, which is able to grab and release it. To this end the RNN was trained using the proprioceptive information collected from the robot. The proprioceptive information consisted of sensorimotor data (i.e. joint positions) and of verbal commands given to the robot according to directions. In the second phase, we tested the ability of the RNN to model mental imagery providing only the auditory stimulus (i.e., the verbal commands) and requiring the network to obtain sensorimotor information from its own outputs.

2. The goal of the second experiment is to test the ability of the RNN to imagine how to accomplish a new task. In this case we had three phases: in the first phase (real training), the RNN was trained to throw front and just to move left and right (with no throw). In the second phase (imagined action), the RNN was asked to imagine its own subsequent sensorimotor state when the throw command is issued together with a side command (left or right). In the final phase (mental training), the input/output series obtained are used for an additional mental training of the RNN. After the first and third phase, experimental tests with the iCub simulator were made to measure the performance of the RNN to control the robot.
A. RNN to model artificial mental imagery

In this experiment we tested the ability of the RNN to recreate its own subsequent sensorimotor state in absence of external stimuli.

In Figure 2, we present a comparison of training and imagined trajectories of learned movements according to the verbal command issued, (a) shows results with the FRONT command, (b) with the BACK command, (c) with the RIGHT command and (d) for the LEFT command. Imagined trajectories are accurate with respect to the ones used to train the robot, only in Figure 3(b) we notice a slight difference between imagined and training positions of the arm. This difference can be attributed to the fact that the BACK command is the only one that does not require the arm to stop early in throwing the object. In other words, the difference is related to the timing of the movement rather than to the accuracy.

Results, presented in Figure 2, show that the RNN is able to recall the correct trajectories of the movement according to the verbal command issued. The trajectories are the sequence of joint positions adopted in the movements.

B. Autonomous learning of new tasks through mental training

This test was conducted to evaluate the ability of the RNN to build its own subsequent sensorimotor states when it is asked to accomplish new tasks not experienced before. In this case the RNN was trained only to throw front and to move right and left (without throwing). To allow the RNN to generalize, training examples were created using an algorithm that randomly chose joint positions not involved in the movement, i.e. when throwing, the torso joint had a fixed position that was randomly chosen. The same was true for arm joints when moving right and left.

Test cases, presented in Figure 4, were composed using two commands (e.g. throw together with right or left). In our experiments we tested two different approaches in the language processing. In this test two commands were computed at the same time, so that input neurons associated with throw and right (or left) were fully activated at the same time with value 1.
Figure 4. Testing cases for autonomous learning with iCub simulator, pictures present the execution of the two composed tasks: throw left and throw right.

Results of the mental training experiment are presented in Figure 5, that show the error of torso and arm joint position respect the ideal ones. The before mental training column presents the results of tests made without additional training, the after mental training column shows results after the simulated mental training, the imagined action only column refers to totally imagined data (i.e. when the RNN predicts its own subsequent input). Comparing results before and after mental training it could be noticed an improvement in precision of dual command execution, this should be accounted to the additional training that helps the RNN to operate in a condition not experienced before.

It should be noticed also that the throw right task has worse performance compared to that of throw left with iCub simulations, but the same result is not achieved in imagined only action mode. This could be mainly explained by the influence of real proprioceptive information coming from robot joints that modifies the ideal behavior expected by the RNN, as evidenced by the comparison between imagined only and the real tests. In fact, we noticed that when a right command is issued the robot torso is initially moved on the left for few timesteps and then it turns right. Since the total time of the movement is due to the arm movement to throw, the initial error for the torso could not be recovered.

Figure 5. Autonomous learning with iCub simulator: error of torso and arm joint positions respect with ideal ones.

IV. CONCLUSION

In this work we create a computational model whose initial results can be used to implement mental imagery processes in an artificial agent. We use motor stimuli coupled with auditory stimuli, in the initial training phase to teach the robot how to execute movements and associate them to verbal instructions. In later stages, the robot is required to recall or “mentally imagine” either the learned trajectory of movements or create a new trajectory from combined commands, exclusively through its comprehension of the linguistic order, without actually executing the movement. Results show that it succeeds in doing so: the same neurons activated when executing the actual movements, are also activated in their absence and this activation is mediated by linguistic stimuli. Results also show that the new data, obtained from “imagination” of new trajectories of combined commands, could be used to enhance the training phase, simulating the process of “mental training”. This replicates, albeit only in part and in a greatly simplified way, growing neuroscientific, theoretical as well as experimental evidence that points to how motor images in the mind recruit the same brain areas involved in executing the particular actions in question, potentially preparing them for action. These results, while still preliminary, also encourage further studies on the computational modeling of motor imagery and its relations with language for the following reasons: They might permit us to better understand the relationships between mental imagery and mental simulations and how they are related to the performance of complex cognitive tasks in the physical world as well as explore how the processing of action and language might be integrated in biological neural systems. This would allow for the design and improvement of language comprehension and processing systems in cognitive robotics research and in linguistically mediated human-machine interactions. Furthermore, they could also be used to improve the skills of humanoid robots for complex sensorimotor tasks, in ways that are similar to those used with real subjects in mental imagery training for sports or rehabilitation purposes. Future work along this line will focus on further differentiating the training phases of the robot into novice and expert levels. The objective would be that of better understanding whether training with imagery has a greater influence on acquired expert skills as opposed to those of novices, as has been found in a number of studies on mental training with elite athletes. Another approach that is in the process of being implemented with the model, instead considers an alternative scenario, one that would in part explore the limits of mental imaging training in cases of physical rehabilitative therapy.

ACKNOWLEDGMENTS

This work was partially supported by the European Commission FP7 Projects: POETICON++ (ICT-288382) within the Cognitive Systems, Interaction, Robotics unit (FP7 ICT Challenge 2), and ROBOT-ERA (ICT-288899) within the ICT for Health, Ageing Well, Inclusion and Governance unit (FP7 ICT Challenge 5)
REFERENCES


