Emergence of communication in RObots through

Sensorimotor and Social Interaction

PROJECT PERIODIC REPORT

Part 3.1 Publishable summary

Project acronym: **ROSSI**
Project title: **Emergence of communication in RObots through Sensorimotor and Social Interaction**
Grant Agreement no: 216125
Date of latest version of Annex I: 8 February 2011

Funding Scheme: Collaborative project/STREP

Date of latest version of Annex I against which the assessment will be made:

Periodic report: 1st ☐ 2nd ☐ 3rd X

Period covered: from 1/03/2010 to 31/10/2011

Project co-ordinator: Prof. Anna Borghi
Alma Mater Studiorum – Università di Bologna
Tel: +39 051 2091838
E-mail: annamaria.borghi@unibo.it
Project website address: http://www.rossiproject.eu

3.1 Publishable summary

Objectives of the ROSSI project

The aim of the ROSSI project was twofold:

- we used embodied models to test embodied theories of cognition and to compare real humans' performances with the ones of simulated organisms.
- given that robots are expected to become very present in unstructured environments, we verified the possibility of communication between humans and robots by investigating how the different sensorimotor capacities they possess can affect the communication. In this respect, we developed a novel approach to sensorimotor grounding of robotic conceptualization and language use.

To explore the extent to which embodied concepts must be shared in order to facilitate communication, we used two robotic platforms with sensorimotor systems structurally similar to human beings. The control mechanisms of these two robots were based on neural mechanisms underlying human concepts and language. In particular, two types of premotor neurons were modelled: canonical neurons and mirror neurons. As to communication, the former mostly underlie the use of nouns while the latter the use of verbs.

Our main objective consisted in providing the two robotics platforms with the mechanisms for the detection of *object affordances* by allowing them to respond to different object visual characteristics with appropriate motor behaviours. The objective was reached successfully as robots were built, that are able to respond to specific visuo-motor demands, analogous to the ones challenging human beings, in proper manners: for instance, they are able to select the correct precision grip, instead of the power one, when a small object was presented.

As a consequence of endowing the simulated systems with a visual and a motor system, the robots became able to develop *concepts grounded on actions*.

On the top of this developed capability, the simulated systems were added with a basic form of language. This was done by referring objects and actions with *verbal labels*. We used nouns (which rely on models regarding the canonical neuron system and affordances), and verbs (which rely instead on models regarding the mirror neuron system).

Third year achievements

During the third year, we carried out behavioural, TMS and brain imaging studies aimed at getting deeper insights into affordances (*for a detailed description see* WP2) and language grounding (*see* WP3); we developed neurocomputational models by synthesising the results of empirical experiments (*see* WP4) and we implemented the models into two physical robotics platforms (*see* WP5).

1. STABLE AND VARIABLE AFFORDANCES

We focused on (a) Consistency, (b) Shape, (c) Size, (e) Weight, (f) Dangerousness

2. ACTION AND FUNCTION IN OBJECTS REPRESENTATION

We tested whether: a) observing objects activates manipulative or functional information; b) observing hands in the presence or absence of an object, and in the presence of novel objects (e.g., a line), activates motor information.

3. MOTOR REPRESENTATION EVOKED BY NOUNS AND ADJECTIVES AND THEIR COMBINATIONS

As in the previous years, we mainly focused on the linguistic category of verbs by conducting different studies on both nouns and adjectives. Finding a modulation of the motor system operated not only by verbs but by nouns and adjectives as well, would support the idea that embodied cognition theories can be generalized to the processing of the different linguistic categories.

4. PROCESSING NOUNS AND VERBS REFERRING TO ABSTRACT CONCEPTS

During this third year, we sought for understanding if and how abstract concepts are grounded on action.

5. SOCIAL DIMENSION

During this third year we further addressed the issue of whether the social framework to which the sentences refer modulates the motor system.

6. EMERGENCE OF VERBS AND NOUNS

During this third year we carried out a brain imaging studies to specifically investigate the pattern of neural activations elicited by verbs and nouns as compared to those of visual actions and visual objects.

Robotic implementation

For robotic implementations, two different approaches can be identified.

1] The first approach planned to combine the existing MNS2 model and the Chain Model, and to extend such combination by linking psychological, neurophysiological, and brain imaging studies. As a result we obtained a comprehensive system level model, intended to work as a framework from which to draw (smaller) focused models, i.e. the *TRoPICALS model* and a *developmental model.* These models were mainly implemented on the simulated iCub platform, with less emphasis on real-world robotic implementation.

2] The second approach planned to use the affordance formalization as a base and to use the empirical results as inspiration to develop functional models geared towards implementation on robots (iCUB and SCHUNK). The models were: (i) a *neuro-computational* model that takes into account the interpretations on the *mirror/canonical* neurons; (**ii**) a *computational model*, based on the *affordance formalization*, outlining how *verb and noun* concepts can be learned/linked to the sensorimotor experiences of the robot; (**iii**) a *learning by demonstration* method, extended to enable online recognition (as well as generation) of demonstrated actions on a robot. i) The *neuro-computational* model, even though it was not implemented on physical robots, provided a good overall big picture on putting together the experimental findings. The second model (ii) was synthesized to enable the formation/linking of verbs and nouns from the interactions of the robot as means to create a shared understanding of verbal concepts between humans and robots. The third model (iii) was synthesized to duplicate the interaction dynamics observed in mirror/canonical neurons through a unified action generation/recognition system. The last two models were successfully implemented on the iCub humanoid robot platform.

AFFORDANCES

iCUB: we tested the final demonstrator in two scenarios.

a. In the first scenario, when the human makes a movement to one of the objects between the human and the robot, the robot needs to recognize the action before it is finished in order to counter-act to it. Using the motion capture system, different actions on both of the two objects are recorded. To form the training set, a 3D grid is defined for action’s starting points. Closed Loop Primitives are used to generate and recognize shown actions. In another experiment an actor performed one of eight behaviors. iCub blinked when it recognized an action; it reacted with the appropriate counter-action in order to bring its hand to the opposite side of the acted object. When the actor brought its hand away from the object, iCub returned to its steady state, making the cycle starting from the beginning.

b. The second scenario regards the affordances learning with and without human interaction, and the use of the learned affordances for human-robot interaction. iCub can successfully learn the affordances of the objects and how they are affected by the presence of a human at the table.

SCHUNK: we used a grasping system that was mapped into human hand as having a robotic version of the thumb, index and middle human fingers. When new objects were presented and visual features extracted, the affordance memory used the new features to output the suitable reaching vector and let the robot to approach and successfully grasp the object. We tested also the color features of the objects.

LANGUAGE AND COMMUNICATION

iCUB: we used interactions which include a human verbally speaking with iCub. The matching effect resulted in the symbolic representation of the observed behaviour, and grounded iCub's concepts into its sensorimotor experience. We studied also how objects can be represented using their affordances in a simplified setting. We demonstrated that the affordances of objects can be used to categorize the objects themselves by using the fact that different objects afford different actions (or similar actions with different effects). Finally we proposed an alternative second approach: the effects obtained through the interaction of the robot are clustered based on the label from a user.

SCHUNK: We modified the Growing Neural Gas Network (RGNG) algorithm (*see* below) to include new dimensions for language. The intensity modulation provided an effective way for the algorithm to assimilate the new tokens (the new added dimensions) within the existing ones. We measured the memory capability to learn name and size tokens, in addition to the verb tokens learned before. The Intensity based algorithm emerged to perform significantly better at predicting tokens from features.

IMPLEMENTATION OF FULL MODEL WITH IMITATION AND FURTHER INTERACTION

iCUB: we used the features and the effect labels acquired from the human (in case there is a human) to learn affordances, or predict learned affordances. We showed that an observed behaviour can be imitated based on its effect using the verb concepts. Finally we demonstrated that with the learned affordances, iCub can get into complex interactions with a human.

SCHUNK: we used a Growing Neural Gas Network to implement an algorithm suitable to model the grasping data. The network was added with a mass parameter (calling then the network as RGNG) which increment each time that a node wins (or decrement otherwise). RGNG was compared to a Self Organizing Map (SOM) being the latter popular for affordance learning robot. Results showed RGNG capable of learning and thus making accurate predictions in far fewer cycles than SOM and the original GNG. The success obtained after had seen that tokens learning is performed better via Intensity Based RGNG Learning, drove AU to test the performance of moves guided by text commands.

Project logo:

Website: www.rossiproject.eu