A ROS architecture for a Cognitive Social Robot

Ignazio Infantino, Agnese Augello, Umberto Maniscalto, Giovanni Pilato, Pietro Storniolo, Filippo Vella
ICAR-CNR, Istituto di Calcolo e Reti ad Alte Prestazioni,
Consiglio Nazionale delle Ricerche, Via U. La Malfa 153, 90146, Palermo, Italy

***Abstract*—The paper presents a framework allowing a robot to socially interact with human beings, sharing with them some basilar cognitive mechanisms. Robust sensing of the environment and people is strongly linked with the cognitive perceptive low level and influences its motivation. Both long-term memory and short-term memory store relevant data to detect and recognize the social context (and social practice), and the human social behavior. Using both internal and external evaluations, the robot learns and improves its social skills, which take into account its physiological and emotional demands (affiliation, competence, certainty). Social interaction is encoded in the cognitive architecture by considering at the same level the human understanding and the robot communicative actions. This is done by impementing a suitable ROS architecture that allow the robot to use the same human interaction channels (both verbal and nonverbal).**

Keywords— Cognitive Architectures, Social Robots, Human-Robot Interaction.

I. Introduction

In the future, social robots will effectively collaborate with people if they will be able to build a real social connection [9][10]. In such a case, both robots and humans have to create a stable and positive relationship (also including an interpersonal influence) based on mutual attentiveness and responsiveness. Humans are fundamentally cognitive emotional beings, and robots have to recognize and interpret affective signals and build suitable cognitive representations that include emotions, motivations, expectations, and also the effects of the internal physical states [6][17][24][25]. Naturally, the environment can strongly influence the social interaction and it often determines the social context and behavior. Then, the robot has to interpret and interact with humans within the correct social practice, switching also among different social contexts and roles through a proactive behavior [13]. In the past, many approaches have simply exploited the causal connections between cognition and emotion using the classical psychology models such as appraisal theory. But to assure a deep understanding and interpretation of the social-emotional displays, we suppose that a whole cognitive framework should inspire the design of the robot software architecture. A suitable cognitive architecture could allow the social robot to develop their socio-cognitive skills within a sort of *Theory of Mind*, (also known as mental perception, social commonsense, folk psychology, social understanding). In fact, the robot needs the ability to recognize, understand, and predict the human behavior regarding the underlying mental states such as beliefs, intents, desires, feeling [12]. Moreover, the artificial cognitive modules could determine a social behavior of the robot that adheres to the people expectations, their reasoning and their manner of acting. In a specular way, given that people apply a theory of mind to understand the robot regarding these mental states as well, the social interaction could be more natural and support the building of a robust social rapport.

II. The Cognitive Architecture

In previous works, we introduced and explained various modules of a cognitive architecture based on emotions and motivations [1][2][4][5][7][14][15]. Such architecture has been employed with success in different domains and experimentations in real environments. It allowed the robot to interact naturally with normal people, to learn by human examples, to improve its performances considering human feedbacks, to show high-level cognitive capabilities such as creativity. In the present paper, we explain the whole architecture by considering all the abstraction levels and dealing with the whole loop perception-reasoning-acting during social interaction. Figure 1 shows the schema of the proposed cognitive architecture for the fully interactive social robot.

A. *Sensing Capabilities and Demands*

The robot is an embodied artificial agent that, firstly, has to perceive itself to evaluate the external environment. Many robotic applications avoid considering such aspect given that the focus is on the execution of a task (often manipulation, and navigation). The state of the robot is monitored to guarantee the successful execution of the sequence of commands and the robot integrity, and alerts cause an immediate stop.

Before reaching such critical conditions, the internal state the robot could be used to modulate its behavior and to obtain a simple self-representation with respect the *extern* useful to build artificial feelings and emotions. In [23] we proposed a somatosensory system that processes different variables of the robot: the temperatures and currents of joints, the battery charge level, the values from the gyroscope, sonars, laser, and other sensors. The proposed model, based on a soft sensor-based approach, allows the robot to own its *roboceptions*, that, such as human sensations and feelings, are the basis for computing a more sophisticated model of emotions. Roboceptions are strictly related to wellness state of the robot and naturally contribute to influence the motivation.

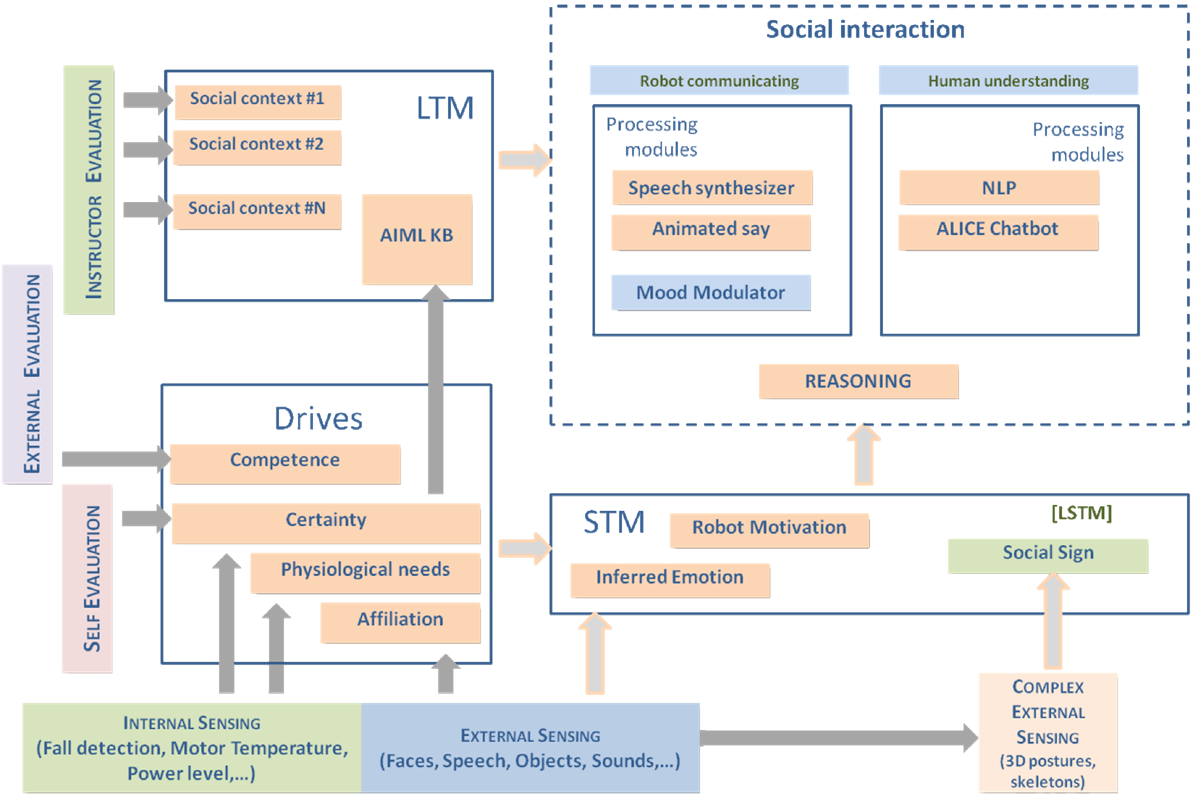

Fig. 1. The Schema of the Cognitive Architecture proposed for an interactive social robot.

The internal sensing is also coupled with some physical feelings caused from external environment: for example, the presence or absence of people, the detection of noise or silence, the presence/absence of obstacles, the level of illumination. External sensing, before being processed to understand and manipulate the environment, causes some physical reaction, as is in living beings. Both internal and external roboceptions determine the physiological state of the robot and constitute one of the four drives (or demands) of the architecture.

For example, in our experimentation of social interaction, we implemented the following roboceptions: pain (from current flowing in the joints), fatigue (from the temperature of the joints), social space or proxemics discomfort (from sonar), caress pleasure (from touch sensors), noise discomfort (from microphones). The algorithm that realizes the soft sensor tries to somehow emulate a biological mechanism by taking into account: the temporal evolution, the memory and the cumulative effects, the latency. A suitable normalization and a weighted sum allows the system to compute various global functions such as well-being, general discomfort, or general distress [23].

The social robot, as a human that someone defines as a social animal, has an innate need to socialize. The affiliation component manages such inclination, and elementary tasks could be automatically executed to satisfy this demand. For example, the robot could look for faces and people, or it could detect and localize human voices and sounds. Each drive has a set of predefined actions to do for increasing the level of satisfaction required by the corresponding demand. Finally, competence and certainty demands consider the capability of the robot to manipulate the environment and its inclination to act. The competence demand is associated to the level of expertise or proficiency with respect an action, a task, or a goal. The robot has to build a collection (a repertoire) of robust and efficient procedures to accomplish various tasks and store them in its Long Term Memory (LTM). The competence demand could be generic or referred to a specific domain and could be measured by counting the number of different procedures available. Competence has to include external evaluation to select good procedures and to discard non satisfactory ones. In the case of an artistic performance, the audience feedback will influence the future executions [3].

The certainty demand deals with the confidence to accomplish a task and it depends from the previous successes or failures with respect some internal evaluation mechanisms (for instance, the expectation arising from the differences with a simulated model). By using suitable weights on the parameters related to the drives, we define the personality of the artificial embodied agent: shy, introverted, gruff, friendly, curious, expansive, sociable, and so on.

B. *Knowledge Representation*

The collection of procedures in LTM that constitute the expertise of the robot arises from learning phases. An instructor drives the learning by explicitly explaining or by showing examples (learning by example or imitation learning) [22]. Moreover, the instructor evaluates the executions to allow the robot to improve over the time. In previous works, we used Interactive Genetic Algorithms to guarantee the evolution of the learning level of the robot [20]. Elementary predefined action modules could be combined (following given rules) to generate various execution planes and they could be selected by a genetic algorithm. To represent the linguistic knowledge of the robot, we use the Artificial Intelligence Markup Language (AIML), that is an XML instance for creating natural language software and used in the implementation of some popular chat-bots. Social procedures (or social practices) could be organized for different social contexts and learned by a *social* instructor. The working memory (or Shor Term Memory, STM) has to detect the features useful to classify and detect the social context. We employ neural networks such as Self Organizing Maps (SOM) to learn to associate contexts to a set of feature values.

In the case of the interaction with a human, the robot detects the social context by recognizing people, objects, facial expression, social signals, and so on. STM, in our architecture, also infers human emotions and compute robot motivation from the drives.

C. *Relevant Issues on Social Interaction*

Depending on the behavior, Breazeal [9][10] proposed four classes of social robots: socially evocative, social interface, social receptive, sociable. Robots belonging to the first two classes, rely on the human tendency to anthropomorphize, and perform just predefined actions perceived as natural, but avoiding a real interaction. Socially receptive robots are considered socially passive, but they can benefit from interaction (e.g., learning skills by imitation). Only at the level of a sociable robot, the artificial embodied agent pro-actively engages with a human to satisfy internal social aims (drives, emotions). Naturally, such a robot requires a deep model of social cognition. In our architecture, we use natural and intuitive communication channels, both to interpret the human behavior and to transfer knowledge to the human. Using natural communication channels includes: the use of natural language and a realistic speech synthesizer; a robust natural speech processing unit; detection and classification of non-verbal cues such as social signals; the generation of non-verbal robot actions to convey emotions and intentions (see the animated say module in the architecture).

A satisfactory verbal interaction [24] requires processing unstructured sentences to infer grammatical and semantic content, searching the appropriate reply in the large repository of knowledge. Considering an active interaction, the robot should drive the evolution of the verbal interaction, looking for acquiring some information from the person such as preferences, desiderata, demands. The robot could ask the person to confirm its understandings or it could require more details to (verbally) react in an appropriate way (for example in the case of ambiguity). To fully understand the human, in the future, the robot has to manage the highest level of a situated language that includes abstract things and concepts both in time and space. Moreover, to assure a robust and not ambiguous social interaction, the robot has to connect the situated language to the physical context performing the so-called symbol grounding, and in some context to perform a meaning negotiation process. At present, we use standard cloud application for robust Natural Language Processing (NLP), and a simple ALICE chat-bot that allow the robot to own a minimalist, stimulus-response language. For instance, the experiment depicted in figure 2 shows a simple social interaction task (based on AIML) to perform a drawing collaborative task. The robot verbally interacts with the user, detects a face, and uses data from a social app to propose a digital collage.

Fig. 2. An example of simple social interaction to perform a drawing collaborative task. The robot verbally interacts with the user, detects a face, uses data from a social app to propose a digital collage.

An important dimension of cognition is the affective/emotional one. The affective dimension is very important in human interaction because it is strongly intertwined with learning, persuasion, and empathy, among many other functions. For the case of speech, affect is marked both in the semantic/pragmatic content as well as in the prosody of speech and the execution of non-verbal movements (head nods, deictic gestures, gaze movements). The Mood Modulator is responsible in the architecture of the affective modulation of the robot communication.

D. *Reasoning and Robot Social Intelligence*

A probabilistic reasoning approach allows the robot to manage uncertainty and the lack of knowledge that is typical in real social interactions. Powerful representations and processing formats are Bayesian Networks (BN) and Markov Decision Processes (MDP), that in the field of knowledge representation and reasoning adopt Bayesian probability for managing evidence and approximation. At present, we exploited a Hidden Markov Model (HMM) based approach, previously used for creative execution of dance [11][19] [21].

Reasoning capabilities are the same required both in generic tasks execution and in social interactions: prediction (often called temporal projection), by inferring what will (probably) happen if the intended course of social action is executed; envisioning, by inferring (all) possible events and effects that will happen if a social practice (as a given sequence of social actions) gets executed in a hypothetical detected social context; diagnosis, by inferring what caused a particular reaction in social practice execution. The social intelligence should allow the robot deciding what is the most appropriate way of interacting with the human in the detected

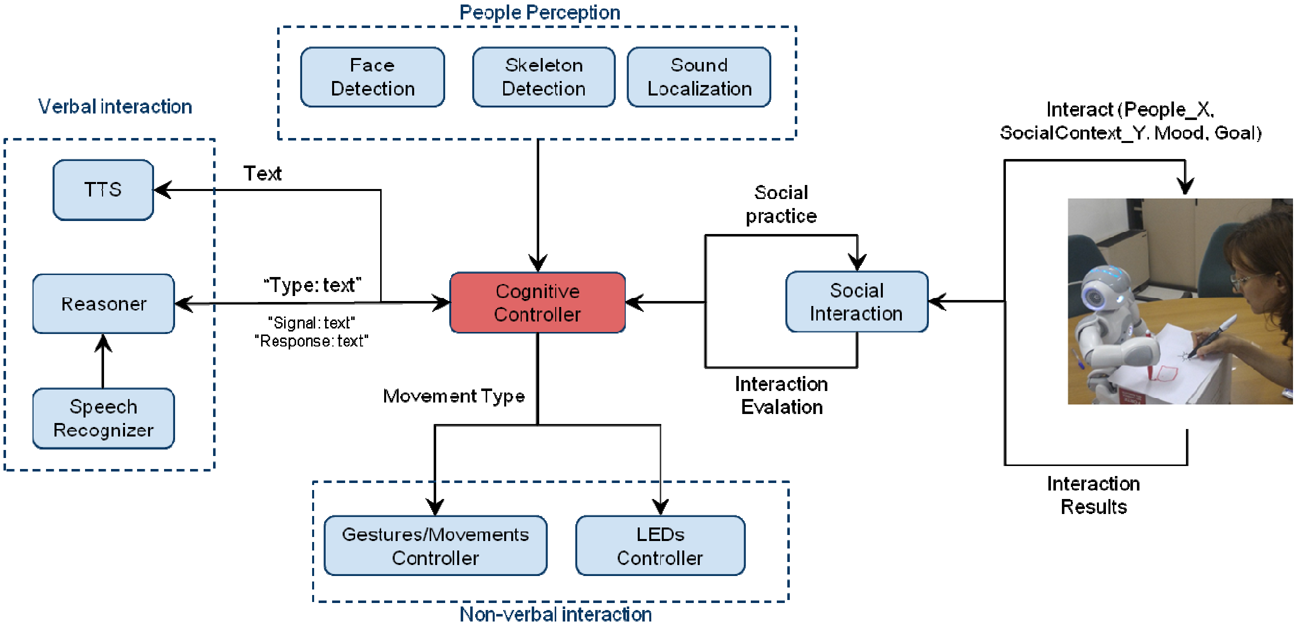

Fig. 3. The ROS architecture for social interaction based on cognitive capzbilities of the robot.

social context, by using classic inference, and heuristic search tools, machine learning methodologies.

A cognitive robot social intelligence is necessary to build a positive and stable social relationship between human beings and the robot, especially in the collaborative task and providing social support in different domains (e.g. education, mental health, physical health, aging). Such a social robot could give relevant social support in different ways: instrumental, informational, emotional, and companionship. Through complex social capabilities, these robots, for instance, could assist people in various ways (by providing information, monitoring performance, incentivizing and sustaining motivation, giving encouragement).

III. The ROS Architecture

We are experimenting with the proposed framework by a suitable ROS architecture presented in Figure 3. The architecture defines main modules and topics that allow the robotic system to assure a cognitive control of the social interaction with a human user. The various scripts belong to the following classes: People Perception, Verbal Interaction, and Non-verbal Interaction. Robot motivation and mood influence the social interaction that requires to adapt to a specific person (People_X), to the detected social context (SocialContext_Y), and to pursue a particular social goal (e.g., reception, affective support, entertainment).

Modules run on a distributed computational architecture, allowing the system to overcome the limitations of the robot computational resources and performances. In fact, the verbal interaction takes advantage of the use of robust speech recognition API such as those provided by Google or similar. Non-verbal interaction uses a deep-learning approach to recognize and classify human action and to produce similar social signs performed by the robot. The core module of the architecture, the Cognitive Controller, is coherent with the framework proposed in previous sections and determines the social behavior. Also, the social interaction requires various levels of evaluation to assure an acceptable behavior and social practice.

The presented system describes a general-purpose architecture for social robots that can exhibit human-like cognitive capabilities. The architecture can integrate various approaches to perform decision making, planning, and reasoning that can take advantage of a simplified cognitive model that explicit motivation, emotion, and feelings.

The aim is to combine perceptions of inner physical states, of the external environment, of peoples with internal and external evaluation mechanisms, allowing the robot to interact with the human showing complex behaviors useful to execute "social practices" and to assure "social support." We are developing a suitable ontology and formalisms to support the knowledge representation based on processing perceptions. In preliminary experiments, since we utilized the humanoid robot Softbank's Pepper, we extend the NaoQi framework based on shared memory (memory) that stores sensors values and events and can represent the working memory. Long-term memory arises from multiple self-organizing maps to classify and represent actions, postures, facial expressions, verbal cues, and all entities that influence the social behavior of the robot. Learning and the evolution of the robot capabilities depend on evaluation based on interactive genetic algorithms.

References

[1] A. Augello, I. Infantino, A. Manfré, G. Pilato, F. Vella, Analyzing and discussing primary creative traits of a robotic artist, Biologically Inspired Cognitive Architectures, vol. 17, 22-31, 2016

[2] A. Augello, I. Infantino, A. Lieto, G. Pilato, R. Rizzo, F. Vella, Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches, Biologically Inspired Cognitive Architectures, vol. 15, pp. 74-86, 2016

[3] A. Augello, I Infantino, A Manfrè, G Pilato, F Vella, A Chella, Creation, and cognition for humanoid live dancing, Robotics and Autonomous Systems, vol. 88, pp107-114, 2016.

[4] A. Augello, I. Infantino, G. Pilato, R. Rizzo, F Vella, Creativity evaluation in a cognitive architecture, Biologically Inspired Cognitive Architectures, vol 11, pp. 29-37, 2015

[5] A. Augello, I. Infantino, G Pilato, R Rizzo, F Vella, Robotic creativity driven by motivation and semantic analysis, Proc. of Semantic Computing (ICSC), 2014 IEEE International Conference on, pp. 285-289, 2014

[6] A. Augello, I Infantino, G Pilato, R Rizzo, F Vella, Combining representational domains for computational creativity, Proc. of the 5th international conference on computational creativity, ICCC 2014, Lubjana, Slovenia, 2014.

[7] A. Augello, I. Infantino, G. Pilato, R. Rizzo, F. Vella, Introducing a creative process on a Cognitive Architecture, Biologically Inspired Cognitive Architectures, Vol. 6, pp. 131-139, Elsevier, 2013.

[8] A. Augello, I Infantino, G Pilato, R Rizzo, F Vella, Binding representational spaces of colors and emotions for creativity, Biologically Inspired Cognitive Architectures, vol. 5, pp. 64-71, 2013

[9] C. Breazeal, Toward sociable robots, Robotics and Autonomous Systems 42 (2003) 167–175.

[10] C. Breazeal, K. Dautenhahn, T., Social Robotics, Springer Handbook of Robotics, Edited by B. Siciliano and O. Khatib, chapter 73, pp. 1935-1960, 2016

[11] G. Città, S. Arnab, A. Augello, M. Gentile, S.I. Zielonka, D. Ifenthaler, I. Infantino, D. La Guardia, A. Manfrè, M. Allegra, Move Your Mind: Creative dancing humanoids as support to STEAM activities, KES-IIMSS 2017, Cognitive Systems and Robotics, Algarve, Portugal, 2017

[12] K. Dautenhahn: Getting to know each other – Artificial social intelligence for autonomous robots, Robotics and Autonomous Systems 16, 333–356 (1995)

[13] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and Autonomous Systems 42 (2003) 143–166

[14] G. Gaglio, I. Infantino, G. Pilato, R. Rizzo, F. Vella (2011). Vision and emotional flow in a cognitive architecture for human-machine interaction. Frontiers in Artificial Intelligence and Applications, Vol. 233: Biologically Inspired Cognitive Architectures 2011, pp. 112 – 117

[15] I. Infantino, Affective Human-Humanoid Interaction Through Cognitive Architecture, The Future of Humanoid Robots - Research and Applications, Dr. Riadh Zaier (Ed.), InTech, 2012

[16] I. Infantino, A. Manfré, U. Maniscalco, Robot Navigation Based on an Artificial Somatosensorial System, First International Early Research Career Enhancement School on Biologically Inspired Cognitive Architectures, pp. 72-77, Springer, 2017

[17] I. Infantino, G. Pilato, R. Rizzo, F. Vella, I Feel Blue: Robots and Humans Sharing Color Representation for Emotional Cognitive Interaction, Biologically Inspired Cognitive Architectures 2012, pp 161-166

[18] I. Infantino, G. Pilato, R. Rizzo, F. Vella, I Feel Blue: Robots and Humans Sharing Color Representation for Emotional Cognitive Interaction, Biologically Inspired Cognitive Architectures 2012, pp 161-166

[19] I. Infantino, A. Augello, A. Manfré, G. Pilato, F. Vella, ROBODANZA: Live Performances of a Creative Dancing Humanoid, Proceedings of the Seventh International Conference on Computational Creativity, Paris, 2016

[20] A. Manfré, A. Augello, G. Pilato, F. Vella, I. Infantino, Exploiting interactive genetic algorithms for creative humanoid dancing, Biologically Inspired Cognitive Architectures, vol. 17, 12-21, 2016

[21] A. Manfrè, I. Infantino, F. Vella, S. Gaglio, An automatic system for humanoid dance creation, Biologically Inspired Cognitive Architectures, vol.15, pp.1-9, 2016

[22] A. Manfrè, I. Infantino, A. Augello, G. Pilato, F. Vella, Learning by demonstration for a dancing robot within a computational creativity framework, IEEE International Conference on Robotic Computing, Taichung, Taiwan, 2017.

[23] U. Maniscalco, I. Infantino, A. Manfrè, Robust mobile robot self-localization by soft sensor paradigm, Robotics and Intelligent Sensors (IRIS), 2017 IEEE International Symposium on, Pages 19-24, IEEE, 2017

[24] Nikolaos Mavridis, A Review of Verbal and Non-Verbal Human-Robot Interactive Communication, https://arxiv.org/pdf/1401.4994.pdf

[25] B. Mutlu, N. Roy, S. Šabanović, Cognitive Human-Robot Interaction, Springer Handbook of Robotics, Edited by B. Siciliano and O. Khatib, chapter 71, pp. 1935-1960, 2016

[26] Mary-Anne Williams, Robot Social Intelligence, in Proc. of Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012