

# Affective Human-Humanoid Interaction Through Cognitive Architecture

Ignazio Infantino

*Istituto di Calcolo e Reti ad Alte Prestazioni, Consiglio Nazionale delle Ricerche  
Italy*

## 1. Introduction

Development of humanoid robots has to address two vital aspects, namely physical appearance and gestures, that will allow the machines to closely resemble humans. Other aspects such as "social" and "emotional" will enable human-machine interaction to be as natural as possible. The field of robotics has long been investigating how effective interaction between humans and autonomous and intelligent mechanical system can be possible (Goodrich & Schultz., 2007). Several distinctive features have been determined depending on whether a robot that acts as an assistant (for example, in the course of a business) or as a companion is required. In the case of humanoid robots, the human appearance and behavior may be very closely linked and integrated if you adopt a cognitive architecture that can take advantage of the natural mechanisms for exchange of information with a human. The robot that cooperates in the execution of an activity would benefit from the execution of its tasks if it had a mechanism that is capable of recognizing and understanding human activity and intention (Kelley et al., 2010), with perhaps the possibility of developing imitation learning by observation mechanisms.

On the other hand, if we consider the robot as a partner, then it plays an important role in sharing the emotional aspects: it is not essential to equip the robot with emotions, but it is important that it can "detect" human emotional states (Malatesta et al. 2009).

The cognitive architectures allow software to deal with problems that require contributions from both the cognitive sciences and robotics, in order to achieve social behavior typical of the human being, which would otherwise be difficult to integrate into traditional systems of artificial intelligence. Several cognitive models of the human mind can find common ground and experimental validation using humanoid agents. For example, if we approach the study of actions and social interactions involving "embodied" agents, the concept of motor resonance investigated in humans may play an important role (Chaminade & Cheng, 2009) to achieve sophisticated, yet simple to implement, imitative behaviors, learning by demonstration, and understanding of the real scene.

In recent years, there is often talk of mirror neurons, which are evidence of the physiological motor resonance at the cellular level with regard to action, action understanding and imitation. But the resonance is applicable in other contexts such as cognitive emotions, the sensations of physical pain, and in various components of the actions of agents interacting socially (Barakova & Lourens, 2009; Fogassi, 2011).

Cognitive models proposed would make the humanoid robot capable of overcoming the so-called "Uncanny Valley of eeriness" (Saygin et al., 2011), by allowing the humanoid is

perceived by human beings not as artificial machine but as a credible social interacting entity. In this sense, the recent experimental data have confirmed the importance of "natural" movements (Saygin et al., 2011) that are expected from the observation of a robot with human features, even if it is a process yet to fully understand, that continually generate predictions about the environment, and compares them with internal states and models. Mirror neurons are assumed to be the neural basis for understanding the goals and intentions (Fogassi 2011), allowing for the prediction of the actions of the individual who is observed, and its intentions. Various studies indicate that they are also involved in the system of empathy, and emotional contagion (Hatfield et al. 1994), explaining that the human tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person.

The classical approach in robotics based on the perception-reasoning-action loop has evolved towards models that unify perception and action, such as the various cognitive theories arising from the Theory of Event Coding (Hommel et al., 2001). Similarly, the objectives are integrated with the intentions, and emotions with reasoning and planning.

An approach that considers the human-robot interaction based on affective computing and cognitive architectures, can address the analysis and reproduction of social processes (and not only) that normally occur between humans, so that a social structure can be created which includes the active presence of a humanoid. Just as a person or group influences the emotions and the behavior of another person or group (Barsade, 2002), the humanoid could play a similar role in owning their own emotional states and behavioral attitudes, and by understanding the affective states to humans to be in "resonance" with them.

The purpose of this chapter is to consider the two aspects, intentions and emotions, simultaneously: discussing and proposing solutions based on cognitive architectures (such as in Infantino et al., 2008) and comparing them against recent literature including areas such as conversational agents (Cerezo et al., 2008).

The structure of the chapter is as follows: firstly, an introduction on the objectives and purposes of the cognitive architectures in robotics will be presented; then a second part on the state of the art methods of detection and recognition of human actions, highlighting those more suitably integrated into an architecture cognitive; a third part on detecting and understanding emotions, and a general overview of effective computing issues; and finally the last part presents an example of architecture that extends on the one presented in (Infantino et al., 2008), and a discussion about possible future developments.

## **2. Cognitive architectures**

To achieve the advanced objective of human-robot interaction, many researchers have developed cognitive systems that consider sensory, motor and learning aspects in a unified manner. Dynamic and adaptive knowledge also needs to be incorporated, basing it on the internal representations that are able to take into account variables contexts, complex actions, goals that may change over time, and capabilities that can extend or enrich themselves through observation and learning. The cognitive architectures represent the infrastructure of an intelligent system that manages, through appropriate knowledge representation, perception, and in general the processes of recognition and categorization, reasoning, planning and decision-making (Langley et al., 2009). In order for the cognitive architecture to be capable of generating behaviors similar to humans, it is important to consider the role of emotions. In this way, reasoning and planning may be influenced by emotional processes and representations as happens in humans. Ideally, this could be

thought of as a representation of emotional states that, in addition to influencing behavior, also helps to manage the detection and recognition of human emotions. Similarly, human intentions may somehow be linked to the expectations and beliefs of the intelligence system. In a wider perspective, the mental capabilities (Vernon et al. 2007) of artificial computational agents can be introduced directly into a cognitive architecture or emerge from the interaction of its components. The approaches presented in the literature are numerous, and range from cognitive testing of theoretical models of the human mind, to robotic architectures based on perceptual-motor components and purely reactive behaviors (see Comparative Table of Cognitive Architectures<sup>1</sup>).

Currently, cognitive architectures have had little impact on real-world applications, and a limited influence in robotics, and the humanoid. The aim and long-term goal is the detailed definition of the Artificial General Intelligence (AGI) (Goertzel, 2007), i.e. the construction of artificial systems that have a skill level equal to that of humans in generic scenarios, or greater than that of the human in certain fields. To understand the potential of existing cognitive architectures and indicate their limits, you must first begin to classify the various proposals presented in the literature. For this purpose, it is useful a taxonomy of cognitive architectures (Vernon et al. 2007; Chong et al., 2007) that identifies three main classes, for example obtained by characteristics such as memory and learning (Duch et al., 2008). In this classification are distinguished symbolic architectures, emerging architectures, and hybrid architectures. In the following, only some architectures are discussed and briefly described, indicating some significant issues that may affect humanoids, and affective-based interactions. At present, there are no cognitive architectures that are strongly oriented to the implementation of embodied social agents, nor even were coded mechanisms to emulate the so-called social intelligence. The representation of the other, the determination of the self, including intentions, desires, emotional states, and social interactions, have not yet had the necessary consideration and have not been investigated approaches that consider them in a unified manner.

The symbolic architectures (referring to a cognitivist approach) are based on an analytical approach of high-level symbols or declarative knowledge. SOAR (State, Operator And Result) is a classic example of an expert rule-based cognitive architecture (Laird et al., 1987). The classic version of SOAR is based on a single long-term memory (storing production-rules), and a single short-term memory (with a symbolic graph structure). In an extended version of the architecture (Laird 2008), in addition to changes on short and long-term memories, was added a module that implements a specific appraisal theory. The intensity of individual appraisals (express either as categorical or numeric values) becomes the intrinsic rewards for reinforcement learning, which significantly speeds learning. (Marinier et al., 2009) presents a unified computational model that combines an abstract cognitive theory of behavior control (PEACTIDM) and a detailed theory of emotion (based on an appraisal theory), integrated in the SOAR cognitive architecture. Existing models that integrate emotion and cognition generally do not fully specify why cognition needs emotion and conversely why emotion needs cognition. Looking ahead, we aim to explore how emotion can be used productively with long-term memory, decision making module, and interactions.

The interaction is a very important aspect that makes possible a direct exchange of information, and may be relevant both for learning to perform intelligent actions. For

---

<sup>1</sup>Biologically Inspired Cognitive Architectures Society -Toward a Comparative Repository of Cognitive Architectures, Models, Tasks and Data. <http://bicasociety.org/cogarch/>

example, under the notion of embodied cognition (Anderson, 2004), an agent acquires its intelligence through interaction with the environment. Among the cognitive architectures, EPIC (Executive Process Control Interactive) focuses his attention on human-machine interaction, aiming to capture the activities of perception, cognition and motion. Through interconnected processors working in parallel are defined patterns of interaction for practical purposes (Kieras & Meyer, 1997).

Finally, among the symbolic architecture, physical agents are relevant in ICARUS (Langley & Choy, 2006), integrated in a cognitive model that manages knowledge that specify the reactive abilities, reactions depending on goals and classes of problems. The architecture consists of several modules that bind in the direction of bottom-up concepts and percepts, and in a top-down manner the goals and abilities. The conceptual memory contains the definition of generic classes of objects and their relationships, and the skill memory stores how to do things.

The emergent architectures are based on networks of processing units that exploit mechanisms of self-organizations and associations. The idea behind this architecture is based on connectionism approach, which provides elementary processing units (processing element PE) arranged in a network that changes its internal state as a result of an interaction. From these interactions, relevant properties emerge, and arise from the memory considered globally or locally organized. Biologically inspired cognitive architectures distribute processing by copying the working of the human brain, and identify functional and anatomical areas correspond to human ones such as the posterior cortex (PC), the frontal cortex (FC), hippocampus (HC). Among these types of architecture, one that is widely used is based on adaptive resonance theory ART (Grossberg, 1987). The ART unifies a number of network designs supporting a myriad of interaction based learning paradigms, and address problems such as pattern recognition and prediction. ART-CogEM models use cognitive-emotional resonances to focus attention on valued goals.

Among the emerging architectures, are also considered models of dynamic systems (Beer 2000, van Gelder & Port, 1996) and models of enactive systems. The first might be more suitable for the development of high-level cognitive functions as intentionality and learning. These dynamic models are derived from the concept that considers the nervous system, body and environment as dynamic models, closely interacting and therefore to be examined simultaneously. This concept also inspired models of enactive systems, but emphasize the principle of self-production and self-development. An example is the architecture of the robot iCub (Sandini et al., 2007), that also includes principles Global Workspace Cognitive Architecture (Shanahan, 2006) and Dynamic Neural Field Architecture (Erlhagen and Bicho, 2006). The underlying assumption is that cognitive processes are entwined with the physical structure of the body and its interaction with the environment, and the cognitive learning is an anticipative skill construction rather than knowledge acquisition.

Hybrid architectures are approaches that combine methods of the previous two classes. The best known of these architectures is ACT-R (Adaptive Components of Rational-thought), which is based on perceptual-motor modules, memory modules, buffers, and pattern matchers. ACT-R (Anderson et al., 2004) process two kinds of representations: declarative and procedural: declarative knowledge is represented in form of chunks, i.e. vector representations of individual properties, each of them accessible from a labeled slot; procedural knowledge is represented in form of productions. Other popular hybrid architectures are: CLARION- The Connectionism Learning Adaptive rule Induction ON-Line (Sun, 2006), LIDA-The Learning Intelligent Distribution Agent (Franklin & Patterson, 2006).

More interesting for the purposes of this chapter is the PSI model (Bartl & Dorner, 1998; Bach et al., 2006) and its architecture that involves explicitly the concepts of emotion and

motivation in cognitive processes. MicroPsi (Bach et al., 2006) is an integrative architecture based on PSI model, has been tested on some practical control applications, and also on simulating artificial agents in a simple virtual world. Similar to LIDA, MicroPsi currently focuses on the lower level aspects of cognitive process, not yet directly handling advanced capabilities like language and abstraction. A variant of MicroPsi framework is included also in CogPrime (Goertzel, B. 2008). This is a multi-representational system, based on a hyper-graphs with uncertain logical relationships and associative relations operating together. Procedures are stored as functional programs; episodes are stored in part as “movies” in a simulation engine.

### 3. Recognition of human activities and intentions

In the wider context of capturing and understanding human behavior (Pantic et al., 2006), it is important to perceive (detect) signals such as facial expressions, body posture, and movements while being able to identify objects and interactions with other components of the environment. The techniques of computer vision and machine learning methodologies enable the gathering and processing of such data in an increasingly accurate and robust way (Kelley et al., 2010). If the system captures the temporal extent of these signals, then it can make predictions and create expectations of their evolution. In this sense, we speak of detecting human intentions, and in a simplified manner, they are related to elementary actions of a human agent (Kelley et al., 2008).

Over the last few years has changed the approach pursued in the field of HCI, shifting the focus on human-centered design for HCI, namely the creation of systems of interaction made for humans and based on models of human behavior (Pantic et al., 2006). The Human-centered design, however, requires thorough analysis and correct processing of all that flows into man-machine communication: the linguistic message, non-linguistic signals of conversation, emotions, attitudes, modes by which information are transmitted, i.e. facial expressions, head movements, non-linguistic vocalizations, movements of hands and body posture, and finally must recognize the context in which information is transmitted. In general, the modeling of human behavior is a challenging task and is based on the various behavioral signals: affective and attitudinal states (e.g. fear, joy, inattention, stress); manipulative behavior (actions used to act on objects environment or self-manipulative actions like biting lips), culture-specific symbols (conventional signs as a wink or a thumbs-up); illustrators actions accompanying the speech, regulators and conversational mediators as who nods the head and smiles.

Systems for the automatic analysis of human behavior should treat all human interaction channels (audio, visual, and tactile), and should analyze both verbal and non verbal signals (words, body gestures, facial expressions and voice, and also physiological reactions). In fact, the human behavioral signals are closely related to affective states, which are conducted by both physiological and using expressions. Due to physiological mechanisms, emotional arousal affects somatic properties such as the size of the pupil, heart rate, sweating, body temperature, respiration rate. These parameters can be easily detected and are objective measures, but often require that the person wearing specific sensors. Such devices in future may be low-cost and miniaturized, distributed in clothing and environment, but which are now unusable on a large scale and in non structured situations. The visual channel that takes into account facial expressions and gestures of the body seems to be relatively more important to human judgment that recognizes and classifies behavioral

states. The human judgment on the observed behavior seems to be more accurate if you consider the face and body as elements of analysis.

A given set of behavioral signals usually does not transmit only one type of message, but can transmit different depending on the context. The context can be completely defined if you find the answers to the following questions: Who, Where, What, How, When and Why (Pantic et al., 2006). These responses disambiguating the situation in which there are both artificial agent that observes and the human being observed.

In the case of human-robot interaction, one of the most important aspects to be explored in the detection of human behavior is the recognition of the intent (Kelley et al., 2008): the problem is to predict the intentions of a person by direct observation of his actions and behaviors. In practice we try to infer the result of a goal-directed mental activity that is not observable, and characterizing precisely the intent. Humans recognize, or otherwise seek to predict the intentions of others, using the result of an innate mechanism to represent, interpret and predict the actions of the other. This mechanism probably is based on taking the perspective of others (Gopnick & Moore, 1994), allowing you to watch and think with eyes and mind of the other.

The interpretation of intentions can anticipate the evolution of the action, and thus capture its temporal dynamic evolution. An approach widely used in statistical classification of systems that evolve over time, is what uses Hidden Markov Model (Duda et al., 2000). The use of HMM in the recognition of intent (emphasizing the prediction) has been suggested in (Tavakkoli et al., 2007), that draws a link between the HMM approach and the theory of the mind.

The recognition of the intent intersects with the recognition of human activity and human behavior. It differs from the recognition of the activity as a predictive component: determining the intentions of an agent, we can actually give an opinion on what we believe are the most likely actions that the agent will perform in the immediate future. The intent can also be clarified or better defined if we recognize the behavior. Again the context is important and how it may serve to disambiguate (Kelley et al., 2008). There are a pairs of actions that may appear identical in every aspect but have different explanations depending on their underlying intentions and the context in which they occur.

Both to understand the behaviors and the intentions, some of the tools necessary to address these problems are developed for the analysis of video sequences and images (Turaga et al., 2008). The aspects of security, monitoring, indexing of archives, led the development of algorithms oriented to the recognition of human activities that can form the basis for the recognition of intentions and behaviors. Starting from the bottom level of processing, the first step is to identify the movements in the scene, to distinguish the background from the rest, to limit the objects of interest, and to monitor changes in time and space. We use then, techniques based on optical flow, segmentation, blob detection, and application of space-time filters on certain features extracted from the scene.

When viewing a scene, the man is able to distinguish the background from the rest, that is, instant by instant, automatically rejects unnecessary information. In this context, a model of attention is necessary to select the relevant parts of the scene correctly. One problem may be, however, that in these regions labeled as background is contained the information that allows for example the recognition of context that allows the disambiguation. Moreover, considering a temporal evolution, what is considered as background in a given instant, may be at the center of attention in successive time instants.

Identified objects in the scene, as well as being associated with a certain spatial location (either 2D, 2D and 1/2, or 3D) and an area or volume of interest, have relations between

them and with the background. So the analysis of the temporal evolution of the scene, should be accompanied with a recognition of relationships (spatial, and semantic) between the various entities involved (the robot itself, humans, actions, objects of interest, components of the background) for the correct interpretation of the context of action. But defining the context in this way, how can we bind the contexts and intentions? There are two possible approaches: the intentions are aware of the contexts, or vice versa the intentions are aware of the contexts (Kelley et al., 2008). In the first case, we ranked every intention carries with it all possible contexts in which it applies, and real-time scenario is not applicable. The second approach, given a context, we should define all the intentions that it may have held (or in a deterministic or probabilistic way). The same kind of reasoning can be done with the behaviors and habits, so think of binding (in the sense of action or sequence of actions to be carried out prototype) with the behaviors.

A model of intention should be composed of two parts (Kelley et al, 2008): a model of activity, which is given for example by a particular HMM, and an associated label. This is the minimum amount of information required to enable a robot to perform disambiguation of context. One could better define the intent, noting a particular sequence of hidden states from the model of activity, and specifying an action to be taken in response. A context model, at a minimum, shall consist of a name or other identifier to distinguish it from other possible contexts in the system, as well as a method to discriminate between intentions. This method may take the form of a set of deterministic rules, or may be a discrete probability distribution defined on the intentions which the context is aware.

There are many sources of contextual information that may be useful to infer the intentions, and perhaps one of the most attractive is to consider the so-called affordances of the object, indicating the actions you can perform on it. It is possible then builds a representation from probabilities of all actions that can be performed on that object. For example, you can use an approach based on natural language (Kelley et al., 2008), building a graph whose vertices are words and a label is the weighed connecting arc indicating the existence of some kind of grammatical relationship. The label indicates the nature of the relationship, and the weight can be proportional to the frequency with which the pair of words exists in that particular relationship. From such a graph, we can calculate the probability to determine the necessary context to interpret an activity. Natural language is a very effective vehicle for expressing the facts of the world, including the affordances of the objects.

If the scene is complex, performance and accuracy can be very poor when you consider all the entities involved. then, can be introduced for example the abstraction of the interaction space, where each agent or object in the scene is represented as a point in a space with a defined distance on it related to the degree of interaction (Kelley et al, 2008). In this case, then consider the physical artificial agent (in our case the humanoid) and its relationship with the space around it, giving more importance to neighboring entities to it and ignore those far away.

#### **4. Detection of human emotions**

Detection of human emotions plays many important roles in facilitating healthy and normal human behavior, such as in planning and deciding what further actions to take, both in interpersonal and social interactions. Currently in the field of human-machine interfaces, systems and devices are now being designed that can recognize, process, or even generate emotions (Cerezo et al., 2008). The "affect recognition" often requires a multidisciplinary and multimodal approach (Zeng et al., 2009), but an important channel that is rich with

information is facial expressiveness (Malatesta et al., 2009). In this context, the problem of expression detection is supported by robust artificial vision techniques. Recognition has proven critical in several aspects: such as in defining basic emotions and expressions, the subjective and cultural variability, and so on.

Consideration must be given the more general context of affect, for which research in psychology has identified three possible models: categorical, dimensional and appraisal-based approach (Grandjean et al., 2008). The first approach is based on the definition of a reduced set of basic emotions, innate and universally recognized. This model is widely used in automatic recognition of emotions, but as well as for human actions and intentions, can be considered more complex models that address a continuous range of affective and emotional states (Gunes et al., 2011). Dimensional models are described by geometric spaces that can use the basic emotions, but represented by a continuous dynamic dimensions such as arousal, valence, expectation, intensity. The appraisal-based approach requires that the emotions are generated by a continuous and recursive evaluation and comparison of an internal state and the state of the outside world (in terms of concerns, needs, and so on). Of course this model is the most complex to achieve the recognition, but is used for the synthesis of virtual agents (Cerezo et al., 2008).

As mentioned previously, most research efforts on the recognition and classification of human emotions (Pantic et al., 2006) focused on a small set of prototype expressions of basic emotions related to analyzing images or video, and analyzing speech. Results reported in the literature indicate that typically performances reach an accuracy from 64% to 98%, but detecting a limited number of basic emotions and involving small groups of human subjects. It is appropriate to identify the limitations of this simplification. For example, if we consider the dimensional models, it becomes important to distinguish the behavior of the various channels of communication of emotion: the visual channel is used to interpret the valence, and arousal seems to be better defined by analyzing audio signals. By introducing a multi-sensory evaluation of emotion, you may have problems of consistency and masking, i.e. that the various communication channels indicate different emotions (Gunes et al., 2011).

Often the emotion recognition systems have aimed to the classification of emotional expressions deduced from static and deliberate, while a challenge is on the recognition of spontaneous emotional expressions (Bartlett, et al. 2005; Bartlett, et. Al. 2006 ; Valstar et al., 2006), i.e. those found in normal social interactions in a continuous manner (surely dependent on context and past history), capable of giving more accurate information about affective state of human involved in a real communication (Zeng et al., 2009).

While the automatic detection of the six basic emotions (including happiness, sadness, fear, anger, disgust and surprise) can be done with reasonably high accuracy, as they are based on universal characteristics which transcend languages and cultures (Ekman, 1994), spontaneous expressions are extremely variable and are produced - by mechanisms not yet fully known - by the person who manifests a behavior (emotional, but also social and cultural), and underlies intentions (conscious or not).

If you look at human communication, some information is related to affective speech, and in particular to the content. Some affective mechanisms of transmission are clear and directly related to linguistics, and other implicit (paralinguistic) signals that may affect especially the way in which words are spoken. You can then use some of the dictionaries that can link the word to the affective content and provide the lexical affinity (Whissell, 1989). In addition, you can analyze the semantic context of the speech to determine more emotional content, or endorse those already detected. The affective messages transmitted through paralinguistic signals, are primarily affected by prosody (Juslin & Scherer, 2005), which may be indicative

of complex states such as anxiety, boredom, and so on. Finally, are also relevant non-linguistic vocalizations such as laughing, crying, sighing, and yawning (Russell et al., 2003). Considering instead the channel visual, emotions arise from the following aspects: facial expressions, movements (actions) facial movements and postures of the body (which may be less susceptible to masking and inconsistency).

Most of the work on the analysis and recognition of emotions is based on the detection of facial expressions, addressing two main approaches (Cohn, 2006; Pantic & Bartlett, 2007): the recognition based on elementary units of facial muscles action (AU), that are part of the coding system of facial expression called the Facial Action coding - FACS (Ekman & Friesen 1977), and recognition based on spatial and temporal characteristics of the face.

FACS is a system used for measuring all visually distinguishable facial movements in terms of atomic actions called Facial Action Unit (AU). The AU is independent of the interpretation, and can be used for any high-level decision-making process, including the recognition of basic emotions (Emotional FACS - EMFACS<sup>2</sup>), the recognition of various emotional states (FACSAID - Facial Action Coding System Affect Interpretation Dictionary<sup>2</sup>), and the recognition of complex psychological states such as pain, depression, etc.. The fact of having a coding, has originated a growing number of studies on spontaneous behavior of the human face based on AU (e.g., Valstar et al., 2006).

The facial expression can also be detected using various pattern recognition approaches based on spatial and temporal characteristics of the face. The features extracted from the face can be geometric shapes such as parts of the face (eyes, mouth, etc.), or location of salient points (the corners of the eyes, mouth, etc.), or facial characteristics based on global appearance and some particular structures, such as wrinkles, bulges, and furrows. Typical examples of geometric feature-based methods are those that face models described as set of reference points (Chang et al., 2006), or characteristic points of the face around the mouth, eyes, eyebrows, nose and chin (Pantic & Patras, 2006), or grids that cover the whole region of the face (Kotsia & Pitas, 2007). The combination of approaches based on geometric features and appearance is likely (eg Tian et al., 2005) the best solution for the design of systems for automatic recognition of facial expressions (Pantic & Patras, 2006). The approaches based on 2D images of course suffer from the problem of the point of view, which can be overcome by considering 3D models of the human face (eg, Hao & Huang, 2008; Soyel & Demirel, 2008, Tsalakanidou & Malassiotis, 2010).

## 5. Integration of a humanoid vision agent in PSI cognitive architecture

SeARCH-In (Sensing-Acting-Reasoning: Computer understands Human Intentions) is an intentional vision framework scheme oriented towards human-humanoid interactions (see figure 1). It extends on the system presented in the previous work (Infantino et al., 2008), improving vision agent and expressiveness of the ontology. Such a system will be able to recognize user faces, to recognize and track human postures by visual perception. The described framework is organized on two modules mapped on the corresponding outputs to obtain intentional perception of faces and intentional perception of human body movements. Moreover a possible integration of an intentional vision agent in the PSI (Bart & Dorner, 1998; Bach et al., 2006) cognitive architecture is proposed, and knowledge management and reasoning is allowed by a suitable OWL-DL ontology.

---

<sup>2</sup><http://face-and-emotion.com/dataface/general/homepage.jsp>

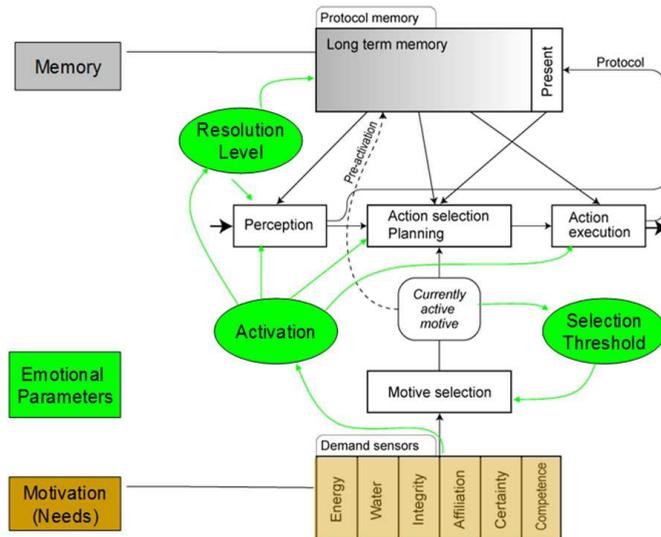


Fig. 1. Cognitive-emotional-motivational schema of the PSI cognitive architecture<sup>3</sup>.

In particular, the ontological knowledge approach is employed for human behavior and expression comprehension, while stored user habits are used for building a semantically meaningful structure for perceiving human wills. A semantic description of user wills is formulated in terms of the symbolic features produced by the intentional vision system. The sequences of symbolic features belonging to a domain specific ontology are employed to infer human wills and to perform suitable actions.

Considering the architecture of PSI (see Figure 1) and the intentional vision agent created by the SEARCH-In framework, you can make some considerations on the perception of the intentions of a human being, the recognition of his identity, the mechanism that triggers of sociality, how memory is used, the symbolic representation of actions and habits, and finally the relationship of the robot's inner emotions and those observed.

The perception that regards the agent is generated from the observation of a human being who acts in an unstructured environment: human face, body, actions, and appearance are the object of humanoid in order to interact with him. The interaction is intended to be based on emotional and affective aspects, on the prediction of intents recalled from the memory and observed previously. Furthermore, the perception concerns, in a secondary way for the moment, the voice and the objects involved in the observed action.

The face and body are the elements analyzed to infer the affective state of the human, and for the recognition of identity. The face is identified in the scene observed by the cameras of the robot using the algorithm of Viola-Jones (Viola & Jones, 2004), and its OpenCV<sup>4</sup> implementation. The implementation of this algorithm is widely used in commercial devices since it is robust, efficient, and allows real-time use. The human body is detected by the Microsoft Kinect device, which is at the moment is external to humanoid, but the data are accessible via the network. From humanoid point of view, the Kinect<sup>5</sup> device is in effect one

<sup>3</sup>Figure available at the link [www.macs.hw.ac.uk/~ruth/psi-refs.html](http://www.macs.hw.ac.uk/~ruth/psi-refs.html) (author: Ruth Aylett)

<sup>4</sup><http://opencv.willowgarage.com/wiki/>

<sup>5</sup><http://en.wikipedia.org/wiki/Kinect>

of its sensor, and the software architecture integrated it as the other sensors. Again, you are using a device that is widely used, and that ensures accurate perceptible results in real time. This sensor produces both a color image of the scene, and a depth map, and the two representations are aligned (registered), allowing you to associate each pixel with the depth estimated by IR laser emitter-detector pair. Through software libraries, it is possible to reconstruct the human posture, through a reconstructed skeleton defined as a set of points in three dimensional space corresponding to the major joints of the human body.

In the region of the image containing the detected face, are run simultaneously two sets of algorithms. The first group is used to capture the facial expression, identifying the position and shape of the main characteristic features of the face: eyes, eyebrows, mouth, and so on. The recognition of the expression is done using the 3D reconstruction (Hao Tang & Huang, 2008), and identifying the differences from a prototype of a neutral expression (see figure 3). The second group, allows the recognition, looking for a match in a database of faces, and using the implementation (API NaoQi, ALFaceDetection module) already available to the NAO humanoid robot (see figure 2).

The PSI model requires that the internal emotional states modulate the perception of the robot, and are conceived as intrinsic aspects of the cognitive model. The emotions of the robots are seen as an emergent property of the process of modulation of the perceptions, behavior, and global cognitive process. In particular, emotions are encoded as configuration settings of cognitive modulators, which influence the pleasure/distress dimension, and on the assessment of the cognitive urges.

The idea of social interaction based on affect recognition and intentions, that is the main theme of this chapter, simply leads to a first practical application of cognitive theory PSI. The detection and recognition of a face meets the need for social interaction that drives the humanoid robot, consistent with the reference theory which deals with social urges or drives, or affiliation. The designed agent includes discrete levels of pleasure/distress: the greatest pleasure is associated with the fact that the robot has recognized an individual, and has in memory the patterns of habitual action (through representations of measured movement parameters, normalized in time and in space, and associated with a label); the lowest level when it detects a not identified face, showing a negative affective state, and a lack of recognition and classification of the observed action. It is possible to implement a simple mechanism of emotional contagion, which executes the recognition of human affective state (limited to an identified human), and tends to set the humanoid on the same mood (positive, neutral, negative). The Nao may indicate his emotional state through the coloring of some leds placed in eyes and ears, and communicates its mood changes by default vocal messages to make the human aware of its status (red is associated with a state of stress, green with neutral state, yellow with euphoria, blue with calm).

The symbolic explicit representation provided by the PSI model requires that the objects, situations, plans are described by a formalism of executable semantic networks, i.e. semantic networks that can change their behaviors via messages, procedures, or changes to the graph. In previous work (Infantino et al., 2008), it has been defined a reference ontology (see figure 3) for the intentional vision agent which together with the semantic network allows for two levels of knowledge representation, increasing the communicative and expressive capabilities.

The working memory, in our example of emotional interaction, simply looks for and identifies human faces, and contains actions for random walk and head movements to allow it to explore space in its vicinity until it finds a human agent to interact with. There is not a world model to compare with the one perceived, even if the reconstructed 3D scene by

depth sensor could be used, and compare it with a similar internal model in order to plane exploration through anticipation in the cognitive architecture. The long-term memory is represented by the collection of usual actions (habits), associated with a certain identity and emotional state, and in relation to certain objects. Again, you might think to introduce simple mechanisms affordances of objects, or introduce a motivational relevance related to the recognition of actions and intentions.



Fig. 2. NAO robot is the humanoid employed to test the agent that integrates SeARCH-In framework and PSI cognitive model.

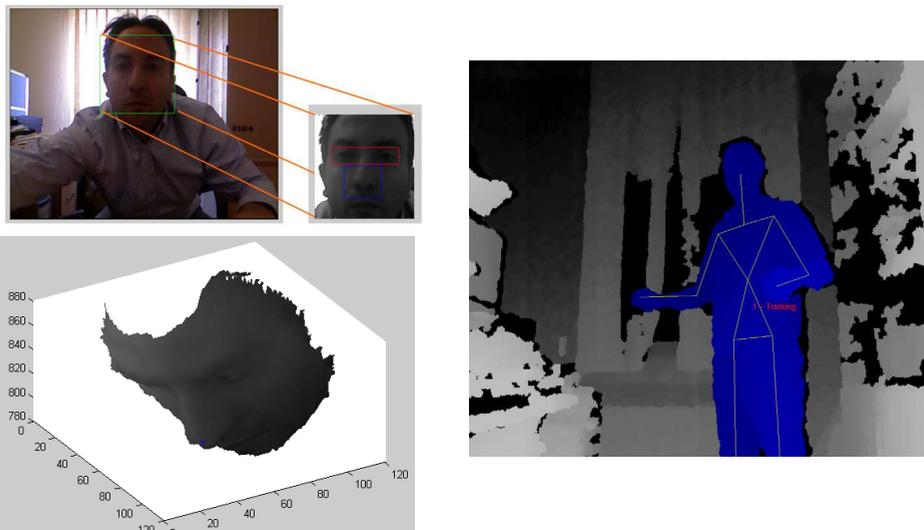


Fig. 3. Example of face and features extraction, 3D reconstruction for expression recognition (on the left), and 3D skeleton of human (on the right).



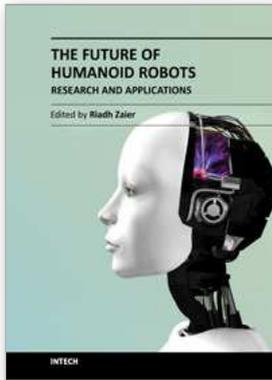
- Bach, J.; Dörner, D., & Vuine, R. (2006). Psi and MicroPsi. A Novel Approach to Modeling Emotion and Cognition in a Cognitive Architecture. Tutorial at ICCM 2006 available at <http://www.macs.hw.ac.uk/EcircusWeb/webContent/>.
- Bachorowski, J., & Fernandez-Dols, J.. (2003). Facial and Vocal Expressions of Emotion. *Ann. Rev. Psychology*, vol. 54, pp. 329-349.
- Barakova, E. I. & Lourens, T. (2009). Mirror neuron framework yields representations for robot interaction, *Neurocomputing*, vol. 72, n. 4-6, Brain Inspired Cognitive Systems (BICS 2006) / Interplay Between Natural and Artificial Computation (IWINAC 2007), January 2009, pp. 895-900. doi: 10.1016/j.neucom.2008.04.057.
- Barsade S.G., (2002). The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior, *Administrative Science Quarterly*, vol. 47, n. 4. (Dec., 2002), pp. 644-675.
- Bartl, C., & Dörner, D. (1998). PSI: A theory of the integration of cognition, emotion and motivation. F. E. Ritter, & R. M. Young (Eds.), *Proceedings of the 2nd European Conference on Cognitive Modelling*, pp. 66-73.
- Bartlett, M.S. ; Littlewort, G. ; Frank, M.; Lainscsek, C.; Fasel, I., & Movellan, J.. (2005). Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '05)*, pp. 568-573.
- Bartlett, M.S. ; Littlewort, G. ; Frank, M.; Lainscsek, C.; Fasel, I., & Movellan, J.. (2006). Fully Automatic Facial Action Recognition in Spontaneous Behavior. In *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR '06)*, pp. 223-230.
- Beer, R. D. (2000) Dynamical approaches to cognitive science, *Trends in Cognitive Sciences*, vol. 4, no. 3, 1 March 2000, pp. 91-99. doi: 10.1016/S1364-6613(99)01440-0.
- Birlo, M & Tapus, A.. (2011) The crucial role of robot self-awareness in HRI, in *Proc. of the 6<sup>th</sup> international conference on Human-robot interaction HRI'11*, ACM. doi>10.1145/1957656.1957688
- Cerezo, E.; Baldassarri, S; Hupont, I. & Seron, F. J. (2008). Affective Embodied Conversational Agents for Natural Interaction, *Affective Computing*, Jimmy Or (Ed.), ISBN: 978-3-902613-23-3, I-Tech Education and Publishing, Available from: [http://www.intechopen.com/articles/show/title/affective\\_embodied\\_conversational\\_agents\\_for\\_natural\\_interaction](http://www.intechopen.com/articles/show/title/affective_embodied_conversational_agents_for_natural_interaction).
- Chaminade, T. & Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics, *Journal of Physiology-Paris*, vol.103, pp. 286-295. doi:10.1016/j.jphysparis.2009.08.011
- Chaminade, T.; Zecca, M.; Blakemore, S.-J.; Takanishi, A.; Frith, C.D. & al. (2010) Brain Response to a Humanoid Robot in Areas Implicated in the Perception of Human Emotional Gestures. *PLoS ONE* 5(7): e11577. doi:10.1371/journal.pone.0011577.
- Chang, Y.; Hu, C.; Feris, R., & Turk, M.. (2006). Manifold Based Analysis of Facial Expression. *Journal of Image and Vision Computing*, vol. 24, no. 6, pp. 605-614.
- Chella, A.; Dindo, H. & Infantino I. (2008). Cognitive approach to goal-level imitation, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, vol. 9, n. 2, pp. 301-318.
- Chong, H. ; Tan, A. & Ng, G.W.. (2007) Integrated cognitive architectures: a survey. *Artificial Intelligence Review*, pp. 103-130
- Cohn, J.F.. (2006). Foundations of Human Computing: Facial Expression and Emotion. In *Proc. Eighth ACM Int'l Conf. Multimodal Interfaces (ICMI '06)*, pp. 233-238.

- Duch, W. ; Oentaryo, R. J. & Pasquier, M.. (2008). Cognitive Architectures: Where do we go from here?. In Proceeding of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference, Pei Wang, Ben Goertzel, and Stan Franklin (Eds.). IOS Press, Amsterdam, The Netherlands, The Netherlands, 122-136.
- Duda, R.; Hart, P. & Stork, D. (2000). Pattern Classification, Wiley-Interscience
- Ekman, P. (1994). Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique," *Psychological Bull.*, vol. 115, no. 2, pp. 268-287.
- Ekman, P. & Friesen, W.. (1977). Facial Action Coding System. Consulting Psychologists Press, 1977.
- Ekman, P.; Friesen, W.V., & Hager, J.C.. (2002). Facial Action Coding System, A Human Face, 2002.
- Epstein, S. L. (1994) For the Right Reasons: The FORR Architecture for Learning in a Skill Domain. *Cognitive Science*, vol. 18, no. 3, pp. 479-511.
- Erlhagen, W. & Bicho, E.. (2006) The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, vol. 3, pp. 36-54.
- Fogassi, L. (2011). The mirror neuron system: How cognitive functions emerge from motor organization, *Journal of Economic Behavior & Organization*, vol. 77, n. 1, Special Issue: Emotions, Natural Selection and Rationality, January 2011, pp. 66-75, doi: 10.1016/j.jebo.2010.04.009.
- Franklin, S. & F. G. Patterson, Jr. (2006). The Lida Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent. Integrated Design and Process Technology, IDPT-2006, San Diego, CA, Society for Design and Process Science.
- Gazzola, V. ; Rizzolatti , G.; B. Wicker, B. & C. Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions, *NeuroImage*, vol. 35, n 4, 1 May 2007, Pages 1674-1684, ISSN 1053-8119, doi: 10.1016/j.neuroimage.2007.02.003.
- Goertzel, B. & Pennachin, C.. (2007). *Artificial General Intelligence*. Springer-Verlag, Berlin, Heidelberg.
- Goertzel, B. (2008). OpenCog Prime: Design for a Thinking Machine. Online wikibook, at <http://opencog.org/wiki/OpenCogPrime>.
- Goodrich, M. A. & Schultz, A (2007). *Human-Robot Interaction : A Survey*, Foundations and Trands in Human-Computer interaction, vol.1, No. 3, 2007, pp. 203-275.
- Gopnick, A. & Moore, A. (1994). "Changing your views: How understanding visual perception can lead to a new theory of mind," in *Children's Early Understanding of Mind*, eds. C. Lewis and P. Mitchell, 157-181. Lawrence Erlbaum
- Grandjean, D.; Sander, D., & Scherer, K. R.. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness & Cognition*, vol. 17, no. 2, pp. 484-495, 2008, *Social Cognition, Emotion, & Self-Consciousness*.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, vol. 11, pp. 23-63.
- Gunes, H. & Piccardi, M. (2009). Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display. *IEEE Transactions on Systems, Man, and Cybernetics Part B, Special Issue on Human Computing* 39(1), 64-84.

- Gunes, H. & Piccardi, M. (2006). A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior. In: Proc. of the 18<sup>th</sup> Intl. Conference on Pattern Recognition (ICPR 2006), vol. 1, pp. 1148–1153 (2006).
- Gunes, H.; Schuller, B.; Pantic, M., & Cowie, R.. (2011). Emotion representation, analysis and synthesis in continuous space: A survey,” in Proc. of IEEE Int. Conf. on Face and Gesture Recognition - 2011.
- Hao, T., & Huang, T.S. (2008). 3D facial expression recognition based on automatically selected features. Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Comp. Society Conf. pp.1-8. doi: 10.1109/CVPRW.2008.4563052
- Hatfield, E; Cacioppo, J. T. & Rapson, R. L. (1994). Emotional contagion. Current Directions, *Psychological Science*, vol. 2, pp. 96-99, Cambridge University Press. doi: 10.1111/j.1467-8721.1993.tb00114.x
- Hommel, B.; Müsseler, J; Aschersleben, G. & Prinz W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, vol. 24, pp. 849-878. doi:10.1017/S0140525X01000103.
- Infantino, I; Lodato, C.; Lopes, S. & Vella, F. (2008). Human-humanoid interaction by an intentional system, In proc. of 8<sup>th</sup> IEEE-RAS International Conference on Humanoids 2008, pp.573-578, 1-3 Dec. 2008. doi: 10.1109/ICHR.2008.4756007.
- Jerritta, S.; Murugappan, M.; Nagarajan, R., & Wan, K.. (2011). Physiological signals based human emotion Recognition: a review. Signal Processing and its Applications (CSPA), IEEE 7th Intl. Colloquium on, pp. 410-415, 4-6 March 2011. doi: 10.1109/CSPA.2011.5759912
- Juslin, P.N., & Scherer, K.R.. (2005). Vocal Expression of Affect. The New Handbook of Methods in Nonverbal Behavior Research, J. Harrigan, R. Rosenthal, and K. Scherer, eds., Oxford Univ. Press.
- Kelley, R.; Tavakkoli, A.; King, C.; Nicolescu, M.; Nicolescu, M. & Bebis, G.. (2008). Understanding human intentions via hidden markov models in autonomous mobile robots. In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction (HRI '08). ACM, New York, NY, USA, 367-374. DOI=10.1145/1349822.1349870
- Kelley, R.; Tavakkoli, A.; King, C.; Nicolescu, M. & Nicolescu, M. (2010). Understanding Activities and Intentions for Human-Robot Interaction, *Human-Robot Interaction*, Daisuke Chugo (Ed.), ISBN: 978-953-307-051-3, InTech, Available from: <http://www.intechopen.com/articles/show/title/understanding-activities-and-intentions-for-human-robot-interaction>.
- Kieras, D. & Meyer, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, vol. 12, pp. 391-438\*MacDorman, K.F. & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, vol. 7, n. 3, pp. 297-337. doi: 10.1075/is.7.3.03mac.
- Kipp, M., & Martin, J.-C. (2009). Gesture and Emotion: Can basic gestural form features discriminate emotions? In proc. of the Intl. Conf. on Affective Computing and Intelligent Interaction 2009, pp.1-8.
- Knox, J. (2010). Response to ‘Emotions in action through the looking glass’, *Journal of Analytical Psychology*, vol. 55, n. 1, pp. 30-34. doi:10.1111/j.1468-5922.2009.01822.x.

- Kotsia I., & Pitas, I. (2007). Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 172-187.
- Laird, J. E. (2008). Extending the Soar Cognitive Architecture. In proc. of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference, Pei Wang, Ben Goertzel, and Stan Franklin (Eds.). IOS Press, Amsterdam, The Netherlands, The Netherlands, pp. 224-235.
- Laird, J. E.; Newell, A. & Rosenbloom, P.S., (1987). "Soar: An architecture for general intelligence," *Artificial Intelligence*, vol. 33, pp. 1-64.
- Langley, P., & Choi, D. (2006). A unified cognitive architecture for physical agents. In proc. of the 21<sup>st</sup> national conference on Artificial intelligence (AAAI'06), vol. 2, pp. 1469-1474.
- Langley, P.; Laird, J. E.; & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10, 141-160.
- Malatesta, L.; Murray, J.; Raouzaoui, A.; Hiolle, A; Cañamero, L. & and Karpouzis, K. (2009). Emotion Modelling and Facial Affect Recognition in Human-Computer and Human-Robot Interaction, *State of the Art in Face Recognition*, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-42-4, I-Tech Education and Publishing, Available from: [http://www.intechopen.com/articles/show/title/emotion\\_modelling\\_and\\_facial\\_affect\\_recognition\\_in\\_human-computer\\_and\\_human-robot\\_interaction](http://www.intechopen.com/articles/show/title/emotion_modelling_and_facial_affect_recognition_in_human-computer_and_human-robot_interaction).
- Marinier R. P. III; Laird, J.E & Lewis, R. L.. (2009). A computational unification of cognitive behavior and emotion, *Cognitive Systems Research*, vol. 10, n. 1, Modeling the Cognitive Antecedents and Consequences of Emotion, March 2009, pp. 48-69, doi: 10.1016/j.cogsys.2008.03.004.
- Pantic, M. & Bartlett, M.S.. (2007). Machine Analysis of Facial Expressions. Face Recognition, K. Delac and M. Grgic, eds., pp. 377-416, I-Tech Education and Publishing Russell, J.A..
- Pantic, M., & Patras, I. (2006). Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments Form Face Profile Image Sequences. *IEEE Trans. Systems, Man, and Cybernetics Part B*, vol. 36, no. 2, pp. 433-449.
- Pantic, M.; Pentland, A.; Nijholt, A. & Huang, T.S. (2006). Human Computing and Machine Understanding of Human Behavior: A Survey, in proc. Of Eighth ACM Int'l Conf. Multimodal Interfaces (ICMI '06), pp. 239-248.
- Russell, J. A.; Bachorowski, J., & Fernández-Dols, J. M. (2003). Facial and Vocal Expressions of Emotion, *Annual Review of Psychology*, vol. 54, pp. 329-349, doi:10.1146/annurev.psych.54.101601.145102
- Salvucci, D. D., & Taatgen, N. A. (2011). Toward a unified view of cognitive control. *Topics in Cognitive Science*, 3, 227-230.
- Sandini, G.; Metta, G. & Vernon, D. (2007). The iCub Cognitive Humanoid Robot: An Open-System Research Platform for Enactive Cognition, in *50 Years of AI*, M. Lungarella et al. (Eds.), Festschrift, LNAI 4850, pp. 359-370, 2007, Springer-Verlag, Heidelberg
- Saygin, A.P.; Chaminade, T.; Ishiguro, H.; Driver, J. & Frith C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc Cogn Affect Neurosci*, first published online April 22, 2011. doi:10.1093/scan/nsr025

- Shanahan, M.. (2006) A cognitive architecture that combines internal simulation with a global workspace, *Consciousness and Cognition*, vol. 15, no. 2, June 2006, pp. 433-449. doi: 10.1016/j.concog.2005.11.005.
- Soyel, H., & Demirel, H.. (2008). 3D facial expression recognition with geometrically localized facial features, *Computer and Information Sciences*, 2008. ISICIS '08. 23<sup>rd</sup> Intl. Symposium on , pp.1-4. doi: 10.1109/ISICIS.2008.4717898
- Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In: Ron Sun (ed.), *Cognition and Multi-Agent Interaction*. Cambridge University Press, New York.
- Tavakkoli, A.; Kelley, R.; King, C.; Nicolescu, M.; Nicolescu, M. & Bebis, G. (2007). "A Vision-Based Architecture for Intent Recognition," Proc. of the International Symposium on Visual Computing, pp. 173-182
- Tax, D., Duin, R. (2004). "Support Vector Data Description." *Machine Learning* 54. pp. 45-66.
- Tian, Y.L.; Kanade, T., & Cohn, J. F.. (2005). Facial Expression Analysis, *Handbook of Face Recognition*, S.Z. Li and A.K. Jain, eds., pp. 247-276, Springer.
- Tistarelli, M. & Grosso, E. (2010). Human Face Analysis: From Identity to Emotion and Intention Recognition, Ethics and Policy of Biometrics, *Lecture Notes in Computer Science*, vol. 6005, pp. 76-88. doi: 10.1007/978-3-642-12595-9\_11.
- Trafton, J. G. ; Cassimatis, N. L. ; Bugajska, M. D. ; Brock, D. P. ; Mintz, F. E. & Schultz, A. C.. (2005) Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 35, n.4, pp. 460-470.
- Tsalakanidou, F. & Malassiotis, S.. (2010). Real-time 2D+3D facial action and expression recognition. *Pattern Recognition*, vol. 43, no. 5, pp. 1763-1775. doi: 10.1016/j.patcog.2009.12.009.
- Turaga, P.; Chellappa, R.; Subrahmanian, V. S. & Udrea, O.. (2008). Machine Recognition of Human Activities: A Survey. *IEEE Treans. On Circuits and Systems for video Technology*, vol18,no. 11, pp. 1473-1488.
- Valstar, M.; Pantic, M; Ambadar, Z. & Cohn, J. F.. (2006). Spontaneous versus Posed Facial Behavior: Automatic Analysis of Brow Actions, in Proc. of Eight Int'l Conf. Multimodal Interfaces (ICMI '06), pp. 162-170.
- van Gelder, T. & Port, R. F.. (1996). It's about time: an overview of the dynamical approach to cognition. In *Mind as motion*, Robert F. Port and Timothy van Gelder (Eds.). Massachusetts Institute of Technology, Cambridge, MA, USA, pp. 1-43.
- Vernon, D.; Metta, G. & Sandini G.. (2007) A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents. *IEEE Transaction on Evolutionary Computation*, vol. 11, n. 2, pp. 151-180.
- Viola, P., & Jones, M. J. (2004) Robust Real-Time Face Detection. *International Journal of Computer Vision*, vol. 57, no 2, pp. 137-154.
- Whissell, C. M.. (1989). The Dictionary of Affect in Language, Emotion: Theory, Research and Experience. The Measurement of Emotions, R. Plutchik and H. Kellerman, eds., vol. 4, pp. 113-131, Academic Press.
- Zeng, Z.; Pantic, M.; Roisman, G.I. & Huang, T.S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.31, no.1, pp.39-58. doi: 10.1109/TPAMI.2008.52



## **The Future of Humanoid Robots - Research and Applications**

Edited by Dr. Riadh Zaier

ISBN 978-953-307-951-6

Hard cover, 300 pages

**Publisher** InTech

**Published online** 20, January, 2012

**Published in print edition** January, 2012

This book provides state of the art scientific and engineering research findings and developments in the field of humanoid robotics and its applications. It is expected that humanoids will change the way we interact with machines, and will have the ability to blend perfectly into an environment already designed for humans. The book contains chapters that aim to discover the future abilities of humanoid robots by presenting a variety of integrated research in various scientific and engineering fields, such as locomotion, perception, adaptive behavior, human-robot interaction, neuroscience and machine learning. The book is designed to be accessible and practical, with an emphasis on useful information to those working in the fields of robotics, cognitive science, artificial intelligence, computational methods and other fields of science directly or indirectly related to the development and usage of future humanoid robots. The editor of the book has extensive R&D experience, patents, and publications in the area of humanoid robotics, and his experience is reflected in editing the content of the book.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ignazio Infantino (2012). Affective Human-Humanoid Interaction Through Cognitive Architecture, *The Future of Humanoid Robots - Research and Applications*, Dr. Riadh Zaier (Ed.), ISBN: 978-953-307-951-6, InTech, Available from: <http://www.intechopen.com/books/the-future-of-humanoid-robots-research-and-applications/affective-human-humanoid-interaction-through-cognitive-architecture>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.