# Can Emotions Enhance the Robot's Cognitive Abilities: a Study in Autonomous HRI with an Emotional Robot

**Ana Tanevska[1,2], Francesco Rea[1], Giulio Sandini[1] and Alessandra Scuitti[1]**

**Abstract.** Cognition represents one of the most important and highly desirable abilities for robots engaged in human-robot interaction (HRI). An ultimate goal is to have a robot that can reason for itself and make decisions about its actions. Several frameworks focusing on different functionalities for autonomous robots have been tried and tested, and a particularly promising venue for cognition focuses on the modelling and development of robotic emotions. This paper presents an on-going project for developing an emotion-based architecture for a humanoid robot with the purpose of having fully autonomous interactions with humans.

## 1 INTRODUCTION

When discussing cognition in the context of robotics, and more precisely for robots engaged in human-robot interaction, the term encompasses several crucial abilities. Echoing the definition of cognitive agents in [1]: a cognitive robot is primarily a robot capable of autonomous interaction with humans, by employing skills such as: environment perception, learning from experience, anticipating outcome of actions, adapting to new circumstances in the environment and ultimately ability to act independently with the purpose of achieving some inner goals. Additionally, depending on the cognitivist or emergent view on cognition, the robot can also be embodied, i.e. cognitively experience the world in a manner that would be unique to its own body and not the same as other robots with the same body and abilities [2].

It becomes evident that implementing full cognition in robots is quite a challenge. However, autonomous robots have been tested in experimental studies for HRI consistently for the past several years. There have been efforts in many different directions towards creating some kind of cognitive architecture or agent to be implemented in robots for interactions with humans. Some of these focus on recognition and employment of joint action [3], others orient towards using cognitive skills in robots that assist humans in achieving goals [4], but there is also an increased amount of studies concerned with emotion-based HRI, both from the aspect of modelling emotions in the robots, and in equipping the robots with functionality for recognizing the emotions of their human peers [5]. Our goals for this project are to see how affect can be modelled and integrated in an interactive robot for cognitive HRI.

[1] Cognitive Robotics and Interaction laboratory. Department of Robotics, Brain and Cognitive Sciences (RBCS), Italian Institute of Technology (IIT), Genoa, Italy. Email: {**ana.tanevska**, **francesco.rea**, **giulio.sandini**, **alessandra.sciutti**}@iit.it.

[2] Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Italy.

## 2 THE ROLE OF EMOTIONS IN COGNITION

From the definition of cognition in the previous section, it becomes evident that often the aim in creating artificial cognitive agents is for them to approach the human cognitive abilities as close as possible.

Even though some researchers opt not to include emotion in their cognitive robots [3-4], emotions (both expressing them and recognizing them in others) still remain a crucial part of cognitive architectures for truly autonomous robots. On one hand, the ability to recognize the expression of emotions in other agents provides the cognitive agent with an information about the state of the environment (i.e. the state of the agents with which is interacting) and can be used to evaluate how well the robot is performing its actions [6-7]. On the other hand, inner emotions constitute the value system in almost all cognitive agents and beings, as emotions are often used both as a motivation and an evaluation mechanism [1].

To summarize, the dual role of emotions in a cognitive architecture can be seen in the two necessary cognitive abilities of *perception* and *motivation*. While a cognitive agent can be non-autonomous and still be capable of learning, anticipation and adaptation, true autonomy is manifested by the presence of inner motivation and autonomous perception of the environment.

### 2.1 Emotions as a motivation and an evaluation mechanism

In cognitive beings (both natural and artificial), motivation represents the main drive of their value system and it is what makes them capable of autonomous action [8]. When it comes to providing a definition of motivation, there is some debate in the community depending on the different outlook (cognitivist or emergent) on cognitive agents. Proponents of cognitivism define motivation as the criteria for selecting a goal and the associated actions leading to that goal; whereas supporters of the emergent theory for cognitive agents interpret motivation as being the encapsulation of the internal value system that modulates the self-regulation and self-development of the agent [9]. Regardless of the finer nuances of the definitions, motivation is the necessary link that enables the agent to act autonomously and make decisions for itself.

There is a general tendency between roboticists to be biased towards human-inspired cognition when we consider cognitive abilities for robots, which is why emotions present themselves as a plausible motivation and value system for the cognitive robot. One way to implement this without delving too deep in the net of bio-inspired design is to consider an architecture for modelling emotions (inspired by the circumplex model [10-11] where we would additionally select between using one or both dimensions) that would provide the robot with a *limited* range of emotions, but additionally implementing those emotions as a value system.

Research done in several HRI studies shows that there are benefits for the interaction even when the robot's emotions don't serve any other purpose from appearance [12-13]. This is particularly the case in child-robot interaction (CRI) studies, where even though the emotions are merely employed as an additional way to make the robot more human-like, the children still exhibit greater ease of interaction and greater engagement than when the robot does not display emotions [14]. However, our goal is to implement this limited range of emotions not only to make the robot more user-friendly, but also to facilitate anticipation, learning and decision-making. A way to modify the existing implementations of emotions would be to make them a part of the anticipation and decision phases in the robot's schema.

To summarize, emotions can be implemented in a cognitive robot as a mechanism for anticipation of future states (in combination with memory), for learning and ultimately for being the motivation or the drive behind decision-making.

## 2.2 Recognition of emotion as a hallmark of autonomy

The previous section highlighted the importance of maintaining and expressing an emotional model for cognitive robots. On the other hand, complete perception of the environment is also a crucial skill for cognitive agents, and this is where the second role of emotion can be seen. When we are discussing a scenario of human-robot interaction, the description of the state of the environment, specifically the state of the human counterpart, is very relevant to the robot. Evidently, as an ultimate goal for HRI we would like to have intelligent robots that would also consider the environment surrounding the human in addition to the human itself, but for the level of HRI scenarios implemented today it is often sufficient to have the robot be able to perceive the state in which the human is.

A perceptive robot that aims at becoming fully autonomous would need to be able to detect the level of engagement, i.e. interest that the user has in the interaction; as well as to be able to identify the possible affect the human might be expressing. Prior results from both HRI and human-computer interaction (HCI) studies show that this two-dimensional evaluation of the state of the user (i.e. evaluation of the affect and engagement) can be sufficient for a balanced interaction that keeps the user engaged and satisfied [6-7][12-13].

This brings us to the point of how the robot should evaluate the state of its human counterpart. There are several different methods used in HRI/HCI studies currently for evaluating and tracking user affect and engagement. Favoured choices include emotion recognition with audio processing (i.e. from speech and emitted sounds) [15-17], body pose estimation and body language tracker (usually done with Kinect sensors) [6][18], detection of micro movements and expressions [19], evaluation by tactile and proximity sensors (so tracking the distance from the robot and the amount of touch) [20], but one of the most reliable and most often implemented is evaluation from facial expressions [21-22].

Evaluating the user's emotions from their facial expression is most commonly done by detecting the intensity of the facial Action Units (AUs) as described in the Facial Action Coding System (FACS) [23]. FACS is based on detecting anatomically based facial actions (i.e. AUs), which are one or few facial muscles that occur individually or in combinations, and can be associated with the expression of certain emotions. There are several open-source software packages for tracking AUs, one of the more commonly used being the OpenFace system [24], which detects the appearance and intensity of 17 facial AUs. These can be used to train models for classifying both the level of engagement and affect of the users.

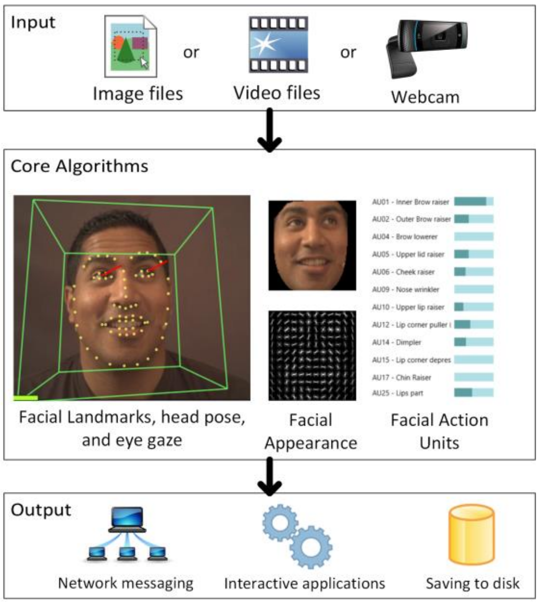


**Figure 1.** The OpenFace Framework

# 3 A COGNITIVE ARCHITECTURE FOR AN EMOTIONALLY-AWARE ROBOT

Having discussed the two roles of emotions in a cognitive agent, as well as some of the ways to implement them as abilities, now we move on to our in-progress architecture for a cognitive robot that would have emotions as its core value system. The cognitive architecture developed for our project has already been partly tested out in the CRI scenario described in [14], which consisted of a study with preschool children in an attempt for autonomous CRI with education modules. The robot was equipped with a self-learning paradigm which enabled it to learn progressively more about the children in terms of which kind of modules each of the children preferred, and in this way keep the interaction novel and adapting.

However, as that project was done on the NAO humanoid robot with a limited scope of modules, the cognitive implementation was more primitive. The project had no functionality for autonomous emotion recognition, and only had *interest* as a one-dimensional variable for the users' (i.e. the children's) moods in the self-learning paradigm. Regardless, the learning paradigm used for the NAO robot proved as effective in enabling the robot to learn and select appropriate actions, so we kept and enhanced it.

## 3.1 Implementation of emotion-driven motivation and learning system

Our aim for this project was to upgrade the existing interaction modules on the iCub humanoid robot with the purpose of achieving cognitive and autonomous HRI. For this, we needed a new architecture that apart from the interaction modules would also include a paradigm to enable the robot to learn on its own and a motivational drive that would provide the robot with the reasoning behind selecting and performing an action.

The first task tackled was the problem of how to make the robot self-learning. A self-learning robot cannot receive any kind of outside teaching or supervision, and instead has to develop and use its own internal mechanism (this in turn constitutes the motivational drive for the robot). Starting from the taxonomy of learning paradigms described in [25], we selected the self-learning paradigm (LI-10) for our architecture. LI-10 is based on using inner emotions (i.e. the robot's own emotions) as an evaluation mechanism instead of relying on outside advice or reinforcement.

The underlying concept of LI-10 is based on the robot iteratively perceiving the state of the environment, considering all possible actions it can undertake, and then selecting the one with the highest probability of affecting its emotional state in a positive direction. The interaction modules for the robot are loosely divided in two groups – modules more oriented towards playful interaction, and modules of a more informative nature. In every iteration the robot can choose whether to carry on with the same module it is already performing, whether to change the module but stay within the same group of modules, or whether to switch to the other group of modules. Evaluating which action is the most beneficial for the robot is carried out via Bayesian nets which express the interdependence between the robot's actions and the changes in the users' emotions.

We mentioned in section 2.1 that it might be better from a computational standpoint to begin with a smaller set of emotions for the robot and eventually expand that set. The three emotion states chosen for our robot are positive (i.e. happy), neutral and negative (sad). As per the principles of LI-10, the robot is directly affected by the changes in the state of the environment, which in our architecture is the level of engagement and affect expressed by the users. More specifically, the robot's mood (expressed as a continuous value between -1.0 and 1.0) varies with the changes of the users' mood – when there is increase in the users' engagement level or expressed affect (or both), the robot's mood also increases, and vice versa.

The self-learning can be observed in the modifications of the Bayesian nets which are unique for each user. The robot begins with two nets containing probability transitions for the connections between users' mood and the robot's action, and iteratively modifies them. If the robot believes that a certain action $A_n$ will cause the user's mood to change from $U_{n1}$ to $U_{n2}$, but in reality the user's mood changes to $U_{n3}$, then the robot modifies the probabilities of transition between those values. More detailed description of how these modifications are carried out can be found in [15].

## 3.2 Emotion recognition from facial expressions

The second important implementation in our architecture was in the module for recognizing emotions. As mentioned in the previous section, the state of the environment is represented by a two-dimensional vector consisting of the values for the users' level of engagement and expressed affect (i.e. arousal and valence [10-11]).

For our architecture we opted to evaluate these metrics from facial expressions, using the OpenFace platform [24] for the detection of AUs and the AM-FED database [26] for training our classifier. The module for recognition and classification of the users' mood has several components. The first component includes the frame-by-frame extraction of AUs and their intensity, as well as other features (like eye gaze, landmark points, head orientation etc.) which are not used for our framework at the moment. After the extraction of the AUs intensity, this data is sent to the second component, which is the module for classifying.

The user's affect is evaluated by the presence of several highly-salient AUs which are unique for separate emotions (i.e. positive affect is marked by the AUs for smiling, cheek raising and crinkling of the eyes, while negative affect is marked by the AUs for brow lowering, nose wrinkling and tightening the lips). If neither of these AUs groups is present in the frame, the user's affect is classified as neutral.

On the other hand, classifying the level of engagement in the interaction is currently done using a Random Forest classifier, which was trained with the AM-FED database which consists of videos depicting people in engaged, neutral or bored states. A comparison was done for the accuracy of several classifiers, including K-nearest neighbours (KNN), Neural Networks (NN), Decision Trees, Support Vector Machines (SVM) and Random Forest. Although KNN had undoubtedly the fastest time for training, it ultimately showed as insufficiently precise, and for the time being we are using Random Forest to classify the engagement. As this is a project still in progress, we do not yet have processed data from testing these modules with users, but the preliminary results from the training show promise for successfully evaluating two-dimensional user moods.

## 3.3 Review of the architecture

In the introduction of this paper we listed the most crucial abilities for a cognitive robot – perception, learning from experience, awareness of the action-reaction principle, ability for anticipation and adaptation, and ultimately ability for independent action with the purpose of achieving some goals. As a final section in this paper, we will give a brief overview of how each one of these skills is implemented in our architecture:

- *Autonomous perception of environment* – for cognitive robots, perception is a specific way of interpreting the environment as experienced by the robot's sensors. In our architecture, the state of the environment is jointly represented by the levels of affect and engagement expressed by the users during the interaction (which can be loosely described as their emotions). The robot perceives these emotions by continuously processing the users facial expressions and classifying them.

- *Ability to learn from experience* – for our robot, learning is the process of acquiring new information about the users as well as learning how the robot's actions affect users' mood. The robot's learning is reflected in the modifications of the probability graphs for each of the users.
- *Anticipating the outcome of actions/events* – similarly as above, the robot's anticipation relies heavily on the probability graphs which represent the robot's knowledge of the world and how its actions might affect the users. The robot 'anticipates' the effect its action might have by calculating which of its actions has the highest probability of affecting the users' mood in a positive manner and performing that action.
- *Adapting to changing circumstances* – adaptation for a cognitive robot includes adjusting to the environment no longer reacting the same way it did before. Adaptation is triggered when an action doesn't have the predicted outcome and is carried out by modifying the probability graphs and adjusting to them for the next iteration.
- *Acting independently to achieve goals* – an autonomous robot acts in a self-driven manner when it is motivated by some inner goals. Our robot was equipped with emotions which were directly influenced by the users' mood – i.e. the more engaged the users were, the happier was the robot. This, in turn, constituted the robot's motivation for selecting an action – perform the action most likely to elicit positive response in the users and in turn most likely to make the robot happier.

# 4 CONCLUSIONS AND FUTURE WORK

Modelling emotions in a cognitive architecture for autonomous robots proved to be a rewarding, if challenging task. In this paper we proposed a complete architecture for evaluating the levels of engagement and interest in the users (i.e. the arousal and valence levels), as well as for providing the robot with a paradigm for self-learning.

Our hopes for our architecture are that it will enable the robot to autonomously perceive the levels of affect and engagement expressed by the users. Additionally, the robot will also be endowed with the ability to learn the users' responses to its performed actions and to anticipate the outcome of its own action. This innovative approach will provide the humanoid robot iCub with the capability to adapt to changing circumstances and to act autonomously to achieve its goals.

For our future work we intend to develop the two novel groups of informative and interactive behaviours for the humanoid robot iCub. Additionally, we plan to evaluate the emerging behaviour in interactive sessions with users. The objective is understand how iCub will adapt specifically to each user, as well as to observe how the users are affected by the behaviour of the robot in ecological interaction.

Our hope for this novel approach to autonomous HRI is to eventually pave the way to long-term, adaptive interaction between humans and robots.


# ACKNOWLEDGEMENTS

This research has been conducted in the framework of the European Project CODEFROR (FP7-PIRSES-2013-612555).


# REFERENCES


[1] Vernon, D. (2014). *Artificial cognitive systems: A primer*. MIT Press.

[2] Chrisley, R. & Ziemke, T. (2002). *Embodiment*. In Encyclopedia of Cognitive Sciences, pp 1102-1108. Macmillan.

[3] Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2016). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence.*

[4] Beer, J., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, *3*(2), 74.

[5] Cañamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural networks*, *18*(4), 445-455.

[6] Vaufreydaz, D., Johal, W., & Combe, C. (2016). Starting engagement detection towards a companion robot using multimodal features. *Robotics and Autonomous Systems*, *75*, 4-16.

[7] Jimenez, F., Yoshikawa, T., Furuhashi, T., & Kanoh, M. (2015). An emotional expression model for educational-support robots. *Journal of Artificial Intelligence and Soft Computing Research*, *5*(1), 51-57.

[8] Merrick, K. E. (2010). A comparative study of value systems for self-motivated exploration and learning by robots. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 119-131.

[9] 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 Transactions on Evolutionary Computation*, *11*(2), 151.

[10] Russell J. A.: A Circumplex Model of Affect. (1980). *Journal of Personality and Social Psychology*, pp. 1161–1178..

[11] Boccanfuso, L., Barney, E., Foster, C., Ahn, Y. A., Chawarska, K., Scassellati, B., & Shic, F. (2016, March). Emotional robot to examine different play patterns and affective responses of children with and without ASD. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 19-26). IEEE.

[12] Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & Mcowan, P. W. (2013). Multimodal affect modeling and recognition for empathic robot companions. *International Journal of Humanoid Robotics*, *10*(01), 1350010.

[13] Soyel, H., & McOwan, P. W. (2013). Towards an affect sensitive interactive companion. *Computers & Electrical Engineering*, *39*(4), 1312-1319.

[14] Tanevska, A. (2016). *Evaluation with emotions in a self-learning robot for interaction with children*. Master's thesis. Faculty of Computer Science and Engineering, Skopje, Macedonia.

[15] Han, K., Yu, D., & Tashev, I. (2014, September). Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech* (pp. 223-227).

[16] Kumar, S. S., & RangaBabu, T. (2015). Emotion and Gender Recognition of Speech Signals Using SVM. *Emotion*, *4*(3).

[17] Pan, Y., Shen, P., & Shen, L. (2012). Speech emotion recognition using support vector machine. *International Journal of Smart Home*, *6*(2), 101-108.

[18] Schwarz, J., Marais, C. C., Leyvand, T., Hudson, S. E., & Mankoff, J. (2014, April). Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 3443-3452). ACM.

[19] Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikäinen, M. (2013, April). A spontaneous micro-expression database: Inducement, collection and baseline. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on* (pp. 1-6). IEEE.

[20] Michalowski, M. P., Sabanovic, S., & Simmons, R. (2006). A spatial model of engagement for a social robot. In *Advanced Motion Control, 2006. 9th IEEE International Workshop on* (pp. 762-767). IEEE.
[21] E. Sariyanidi, H. Gunes, and A. Cavallaro. (2014). Automatic analysis of facial affect: A survey of registration, representation and recognition. IEEE TPAMI.
[22] F. De la Torre and J. F. Cohn. (2011). Facial Expression Analysis. In Guide to Visual Analysis of Humans: Looking at People.
[23] P. Ekman and W. V. Friesen. (1977). Manual for the Facial Action Coding System. Palo Alto: Consulting Psychologists Press.
[24] Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on* (pp. 1-10). IEEE.
[25] Ackovska N. (2010). Taxonomy of Learning Agents. In *Didactical Modeling Annals* Vol 4:Year 2010/2011, Institute of Mathematics and Informatics, Bulgarian Academy of Science.
[26] McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J., & Picard, R. (2013). Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 881-888).