



HAL
open science

Towards an Affective Cognitive Architecture for Human-Robot Interaction for the iCub Robot

Ana Tanevska, Francesco Rea, Giulio Sandini, Alessandra Sciutti

► To cite this version:

Ana Tanevska, Francesco Rea, Giulio Sandini, Alessandra Sciutti. Towards an Affective Cognitive Architecture for Human-Robot Interaction for the iCub Robot. 1st Workshop on “Behavior, Emotion and Representation: Building Blocks of Interaction”, Oct 2017, Bielefeld, Germany. hal-01615491

HAL Id: hal-01615491

<https://inria.hal.science/hal-01615491>

Submitted on 12 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards an Affective Cognitive Architecture for Human-Robot Interaction for the iCub Robot

Ana Tanevska

Francesco Rea

Giulio Sandini

Alessandra Sciutti

Department of Robotics, Brain and Cognitive Science
Italian Institute of Technology
Genoa, Italy

{ana.tanevska,francesco.rea,giulio.sandini,alessandra.sciutti}@iit.it

ABSTRACT

The ability to reason for itself and make decisions for its own actions is a highly desirable ability for robots engaged in human-robot interaction (HRI). By being able to perceive the environment, select the optimal action w.r.t. the state of the environment and then observe and learn from the reactions that occur, a robot can be said to be fully autonomous in the interaction. There have been several different types of frameworks focusing on cognitive functionalities for autonomous robots, and one particularly promising venue for cognition focuses on the affect-driven cognitive frameworks. This paper presents an ongoing study focused on developing an affect-based architecture for the humanoid robot iCub with the purpose of having fully autonomous interactions with humans.

Author Keywords

affective computing; human-robot interaction; humanoid robots; cognitive architecture; cognitive interaction

ACM Classification Keywords

I.2.9 Robotics

INTRODUCTION

Having robots be able to autonomously interact with humans, in a context where they would have to continuously adapt to novel and unstructured scenarios is considered as a significant milestone in HRI. To reach this kind of autonomous behaviour in social robots it becomes necessary for some level of cognition to be implemented in the robot's architecture.

In general, cognition in both natural and artificial agents is seen as the collection of several crucial abilities that enable the autonomous behaviour of the system. Going by the definition of cognition from Vernon [30], a cognitive system or agent has to be capable of: environment perception, learning

from experience, anticipating the outcome of performed actions, adapting to changing circumstances in the environment and ultimately being able to act independently driven by its own motivation. For robots engaged in social HRI, the implementation of some (or all) of these cognitive abilities has been demonstrated in the context of autonomous social interaction [18][2], in studies aimed at the employment of joint action [15] as well as in studies where the robot is assisting the humans in achieving their goals [5].

In our study, we are working in the context of social interaction generalized to fit different settings, where the robot would have interact with people autonomously, fluidly and above all with some level of intelligent behavior. Our aim is for the social robot to have cognitive skills resembling those of humans [6]. Drawing inspiration from people, we propose to endow the robot with the ability to perceive the state of the person it is interacting with and to autonomously decide how to act, guided by an affect-driven motivational drive. There is some debate present on the question of whether affect and emotions in general play a vital part in the human cognitive architecture or whether they are separate from cognition[9]. This latter view can be evidenced in several studies for cognition in artificial systems where emotions are not included in the proposed cognitive architectures [15] [5].

Nonetheless, recently there have been researchers showing that emotions represent a crucial part of the human-inspired cognitive architectures for autonomous robots [29][11]. Although emotional expressiveness can be implemented as a component in different functionalities, there are two main areas where it can serve a crucial part in the cognitive architecture - by representing an important perceptual input for the person in the HRI scenario, and by serving as an internal motivation and evaluation system for the robot. The ability to recognize the expression of emotions in other beings is providing information about the state of the environment after each iteration and can be used to evaluate how well the robot is performing [29][14][7]. There are several different methods used in HRI/HCI studies for evaluating and tracking the affective state of the person in the interaction [18], including emotion recognition from speech [13][23], body pose [29] [26] as well as distance from the robot and amount of touch [20]. However, one of the more reliable and least intrusive method is the analysis of facial features and gaze [25][10][21].

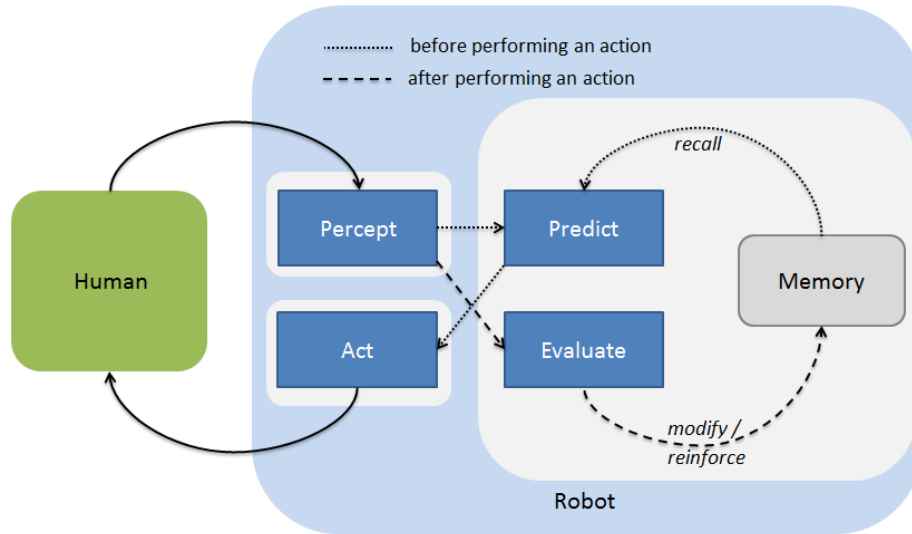


Figure 1: The layout of our architecture.

On the other hand, emotions, if implemented in some pseudo form, can also constitute a value system for the cognitive robot, as emotions are often used both as motivation and evaluation mechanisms [30]. By implementing a (limited) range of pseudo emotions for the robot (eg. happy, sad, neutral, surprised etc), it is possible to make them a part of the anticipation and decision phases in the robot’s schema, supporting learning and adaptation. There are already several examples of artificial cognitive architectures exploiting emotions as inner motivation, such as the FATiMA architecture equipped with modules for appraisal and behaviour [11], the WASABI architecture dealing with implementation of both primary and secondary emotions [4], the CogAff architectural schema (and its upgrade to an actual cognitive architecture - H-CogAff) with reflective, deliberative and reactive processes [27] and the “bidirectional grounding”-based cognitive schema described in [16].

That as may be, combining these two affect-based functionalities in robots is not yet a widespread practice. There have been studies on empathic HRI, e.g. with the iCat robot[8], however, there aren’t many studies which attempt fully autonomous affect-based HRI with more elaborate humanoid platforms. The adoption of such a platform would entail additional challenges, such as the heightened human expectations for the robot’s performance, as well as the complexity of the control of the platform.

The study described in this paper presents our work on developing a cognitive, self-learning architecture for social interaction for the iCub robot. The rest of the paper is organized as follows: *Our Architecture* presents the structure of our cognitive framework and the modules (shown in Figure 1). This is followed by *Methods*, which covers the implementations of the modules in the respective subsections *Perception Module* and *Learning Module*. Finally, in *Results* and *Conclusion* we

present our results from testing the perception and learning components and we discuss our next steps in the study.

OUR ARCHITECTURE

The envisioned goal of our study is to develop a cognitive architecture for the iCub humanoid robot which would provide the iCub with the primary supportive functionalities necessary for autonomous HRI. The main concept is a framework which would allow the robot to perceive and evaluate the affective state of the persons with which it’s interacting and make an intelligent decision when selecting and performing actions. This base framework, once established, can later be generalized to fit many different contexts of HRI social, educational, joint-action, assistive, etc - depending on the types of actions which would be implemented. Our intention is instead of focusing on developing a framework for a narrow set of behaviours, to set the groundwork for a learning and decision-making platform which could be utilized for any kind of interaction between a human and an intelligent robot.

However, this base framework cannot consist solely out of the standard action-perception loop, as it is insufficient to cover the full range of cognitive functionalities for the intelligent robots. To satisfy the abovementioned functionalities necessary for cognition, there is the need for additional components which would be responsible for learning and intelligence, for forming a connection between the robot’s action and the person’s reaction, as well as memorizing the emergent patterns of behavior and response for each person.

The closed loop then, needs to be extended by implementing a basic form of memory, and then adding either a reinforcement mechanism applied after the executed action, which provides feedback to the robot regarding its last action or series of actions, or an advice mechanism, which is implemented after the perception phase and provides the recommended course of action based on the perceived state. In supervised learning

scenarios, this is done via a teacher who provides either the reinforcement or the advice[1]. In self-learning frameworks (as autonomous robots would preferably have), we would like to have both implemented and functioning in the extended closed loop.

In the previous section, we mentioned that the set of primary requirements for cognitive agents, apart from perception and action also included adaptation, anticipation, learning and motivation. For our architecture we propose a motivational system based on affect, where the robot’s emotional state changes for the better or worse depending on the perceived change in the person’s affective state. The knowledge of how each person’s state changes over time and series of interaction needs to be stored in the robot’s memory, and updated whenever the robot learns some new information about the behaviour of the person.

Thus the framework for interaction would take the format of this kind of loop:

- Perceiving the state of the human
- Before performing an action
 - Predicting which action would be most beneficial for robot and human (by improving their affective states)
 - Performing the most beneficial action
- After performing an action
 - Evaluating from perceptual input if the person’s reaction is as predicted
 - Modifying the belief values if wrong, reinforcing them if right

There are three distinct components functioning separately that needed to be developed for our architecture - the perception, internal evaluation/self-learning and action components. The perception component was tasked with obtaining the visual feedback from the human peer, and processing the visual signal into information about the affective state of the person. The evaluation component is the one in charge of anticipating the outcome of the future actions, as well as learning post-action whether its beliefs were true. The action component is the one in charge of performing the action selected as the most potent one.

METHODS

Perception Module

The perception component of our architecture is tasked at the moment with processing just the visual input, although there is consideration for expanding this with spatial or tactile sensor data further on. More specifically, we are processing the monocular information from one of iCub’s cameras about the facial features of the person. The monocular input was selected since at the moment there is no requirement for stereo vision in the framework, and since we wanted to have as fast as possible perception module for real-time analysis of the face during at interaction.

The perception module was firstly implemented on a Windows machine so as to fully test its functionalities before porting it to the iCub. For the processing part we started with some of the functionalities integrated in the OpenFace framework developed by T. Baltrusaitis [3]. OpenFace is an open-source architecture which offers the analysis of images (single image or a series), video or camera feed, and the extraction of all the relevant information about the facial landmarks, head orientation, facial expressions, gaze direction, etc (as shown in Figure 2). However, for the scope of our perception module we just required processing the camera feed and extracting the facial landmarks, so we started with their architecture for detecting the facial features and adapted it further to the iCub software.

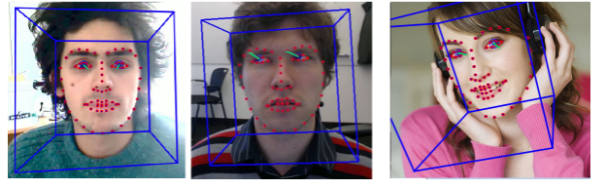


Figure 2: Output from the OpenFace modules.

The perception module consists of several sub-modules. The first processing includes the analysis of the scene from the camera feed and the attempt to detect a face in it. Following the detection of the face, which is done by using the dlib library and its frontal face detector, the landmarks on the human face (shown in Figure 3) were extracted. From the 68 facial landmarks the most important factors for expression analysis can be extracted the facial action units (AUs). The facial action units are anatomically based facial actions (e.g. lifting/lowering eyebrows, smiling, opening mouth, wrinkling nose, etc), which can be associated with the expression of certain emotions (see Figure 4) [12].

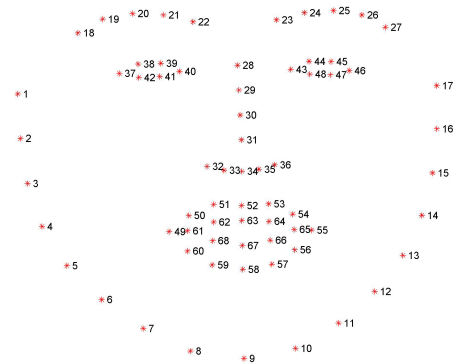


Figure 3: The 68 facial landmarks extracted from a frontal face view.

The extraction of the AUs was done using two phases: extracting the appearance of the AUs by using linear-kernel SVM and coding the appearance or non-appearance of a certain AU as 1 or 0; and extracting the intensity of the AUs

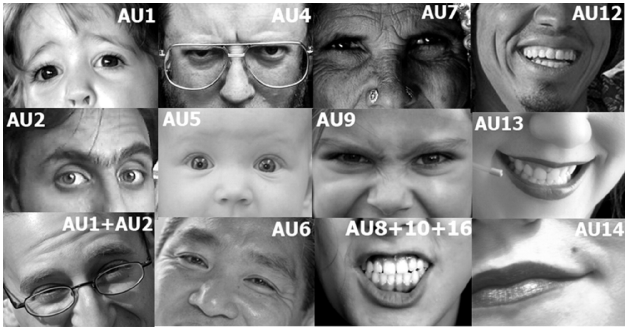


Figure 4: Some of the AUs extracted from facial landmarks.

by using a linear-kernel SVR, where the value given for each AU varies continuously between 0 and 5. Groups of these AUs can be highly exclusive and connected to separate affective states (i.e. anger, happiness, sadness, surprise, etc)[17], which made it very simple to program the affect detection module depending on the summary intensity value of the AUs at a given frame. This has the additional benefit of not requiring the training of a classifier or regressor for detecting the facial expressions, as they have been shown to be linked with the activation of specific subsets of AUs.

We also evaluated whether the same module could be used for distinguishing the engagement state of the person (i.e. whether the person is interested or bored in the interaction). However, unlike the detection and classification of the affective states, which are very salient and can be detected without the need of a trained classifier but just by evaluating and comparing values, to assess the engagement of the persons we needed a classifier.

For the training and testing of the classifier we used the videos of the AMFED database [19]. The database consists of videos of people responding to advertisements, annotated by the self-described reaction of the users about how engaged they felt during the advertisement and if they would watch it again. There were a total of 235 videos, with average duration of 55 seconds. The set of videos was first processed and the intensities of the expressive AUs were extracted for each of the videos, excluding only the first 100 frames which depicted the people getting ready to play the advertisement. The processed AU data was then used to train several classifiers : K-nearest neighbours (KNN), ADA boost, Neural Networks (NN), Decision Trees (DT) and Random Forest (RF). The training was done using the sci-kit library for python and its implementations of the abovementioned classifiers, using the default provided parameters[24].

Learning Module

The learning module was the component providing the link between perception and action. In the previous section we mentioned that the functionalities that needed to be covered were motivation, anticipation, learning and adaptation. We addressed this issue in three steps. Firstly, by defining that every change (be it positive or negative) in the person's affective state impacts the robot's emotions. Next, by defining that the internal goal of the robot is to maximize its own happiness,

by making sure the person's affective state changes for the better. And finally, by implementing a basic system responsible of storing the robot's knowledge about what each person prefers in the interaction, and also responsible of learning and adapting w.r.t. to the observed behavior of each person following a robot's action.

The value system used for these functionalities was implemented in the form of Markov chains of transitional probabilities, with a separate graph for each of the robot's behaviors, and a separate group of graphs for each person separately. A sample graph is depicted in Figure 5. The values of the transitions are the probabilities with which a person's state could change from state A to state B, by performing the selected action.

The scenario covered by this graph shows the values given a robot's behavior of play and the tracked states of neutral, interested and bored for the person. This was extracted from our previous study on affect-based autonomous framework, where the robot could exhibit two different typologies of behavior (i.e. engaging the users to learn something or to play with the robot), and was tracking the engagement state of its human peers [28].

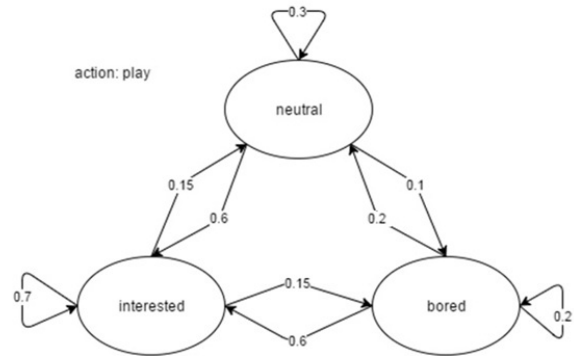


Figure 5: Example of a Markov chain in the learning module.

The process that happens in between perceiving the state of the person and selecting an optimal action is the following: for each of the possible actions (or behaviours) the robot can perform, it needs to calculate which one has the highest probability to improve the state of the person. For instance, if the perception module has evaluated the person is now sad, the robot needs to evaluate which action has the greatest probability to make the robot feel better by improving the state of the person.

Evaluating these probabilities is done by utilizing the Markov chains to calculate the conditional probabilities about the change in the affective states of the robot and the person (see equations below).

$$P(r1 = m1) = \sum_{c1} P(r1 = m1 | r0, u1)P(u1)$$

$$P(u1 = n1) = P(u1 = n1 | u0, a1)$$

Shown in the equations are the calculations for the probabilities of the robot state $r1$ changing to the value $m1$ depending on the affective state of the person $u1$, as well as the probability for the state of the person $u1$ entering the new affective state $n1$, for a given action of the robot $r1$.

This can be further extended if the perceptual input is two-dimensional, i.e. if the robot needs to consider the changes in both the level of engagement of the user and their affective state. Tracking both dimensions can be beneficial in HRI, if we consider that not only do we want to know whether the person is feeling happy or sad or angry, but also whether those emotions are related to the interaction with the robot, or are caused by external factor not related to the robot's behaviour.

There are two methods of tracking the changes in a two-dimensional space: either by noting the combination of the changes (e.g. increased affect + decreased engagement) or by giving preference to one of the dimensions (i.e. whether it's more important to have engaged but angry users, or slightly bored but happy). The decision which way would be the more beneficial one depends strongly on the context of the HRI, which is to say on the types of behaviours exhibited by the robot and the scene of the HRI.

RESULTS AND DISCUSSION

While the system described in this paper is still a work-in-progress, we already have obtained some results from the implementation and testing of the components in the perception module, as well as previous results supporting the integration of the learning module.

Perception Module

Affect detection

The affect detection module was tested successfully in several demos in the form of an emotional mimicry module. This module was executed in parallel with the face tracking module for the iCub, and it presented a scenario where the iCub was following the face of the user, while simultaneously also mimicking the facial expressions the user was showing (see Figure 6).

This was done by comparing the intensity values of the groups of AUs and evaluating the emotional state the user was expressing; and then using the predefined combinations of the LEDs in the face of the iCub to express the same emotional state. The module was tested during demos in our laboratory with both adults and children present, to quite noticeable interest from the audience, and it has also been implemented as a part of a greater composite demos.

Currently, we are also looking into more formal, quantitative testing for the affect detection component with the purpose of obtaining a benchmark for its accuracy, as all of the evaluation so far has been of a more qualitative nature.

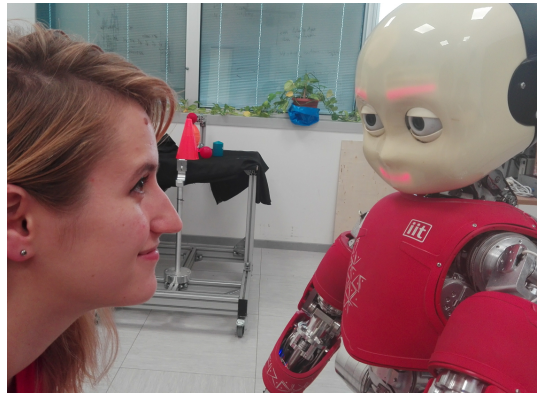


Figure 6: Examples of the mimicry module in action.

Engagement detection

Out of the five classifiers trained on the AM-FED videos, the highest accuracy and efficiency in training showed the Random Forest classifier, implemented with constraints of 10 trees per forest and maximum depth of level 5. Table 1 shows the training time in seconds, and the results from the 5-fold cross-validation for each classifier. Considering the comparison in difficulties for classifying engagement over classifying affect, as well as the fact that recent state-of-the-art accuracy for classifying facial expressions ranges around 75.2%, our result of 69.5% can still be considered satisfactory. However, the situations in which a robot has to assess the engagement of its human peers would be quite different regarding the display of the engagement from the context in which the AM-FED videos were recorded. To evaluate whether also natural interaction with iCub can evoke engagement similarly readable from facial expressions, we processed recordings of an HRI experiment with the iCub, covering the part of the experiment where iCub is explaining the procedure for the experiment to the people, thus engaging them in direct interaction [22].

	KNN	ADA	DT	RF	NN
Accuracy	59.6%	66.8%	68.1%	69.5%	63.7 %
Train time	7.48 s	41.59 s	2.06 s	1.85 s	141.47 s

Table 1: Results from 5-fold cross-validation on the AM-FED database

To obtain a benchmark of the level of engagement exhibited by the people, we asked four raters (2 females, 2 males; avg age: 28years) to watch the same videos and continuously rate the degree of engagement by pressing a key (either 0: disengaged, 1: neutral or 2: engaged). A subject in the video was labeled as engaged for a rater if the rating was 2 for the majority of its duration, disengaged if the rating was 0 most of the time and neutral if it was 1 for most of its duration. For the final calculations, we opted for a binary labeling (engaged vs. disengaged) by accumulating neutral votes to the label containing the higher number of raters and then picking the most voted category as label for the video.

We used the rating from the human judges to obtain the degree of inter-rater reliability and calculate the Fleiss kappa value. However, the results showed a low kappa value of 0.569892 (with a standard error of 0.129099), which falls short of the value of 0.61+, required for detecting "substantial agreement" between raters. This showed that unlike with affective expressions, evaluating the level of engagement in people is difficult even for human judges. Because of that, if we plan on using the facial features to evaluate the engagement level via a classifier, we will need to train it with a more context-relevant dataset.

Learning Module

As of now, the learning component has not yet been tested in an HRI experiment with the current setup on the iCub robot, since its functioning is strongly dependent on the integration with a fully developed action module. However, the same framework for the learning module was already tested out in our previous study, in a child-robot scenario with the NAO humanoid robot [28].

The study involved three pre-school children engaging in playful or educational interaction with the NAO robot, and the framework proved as capable of learning each of the children's preferences during the sessions of interaction. Figure 6 shows the adjustment learned for each of the children after just three sessions of interaction.

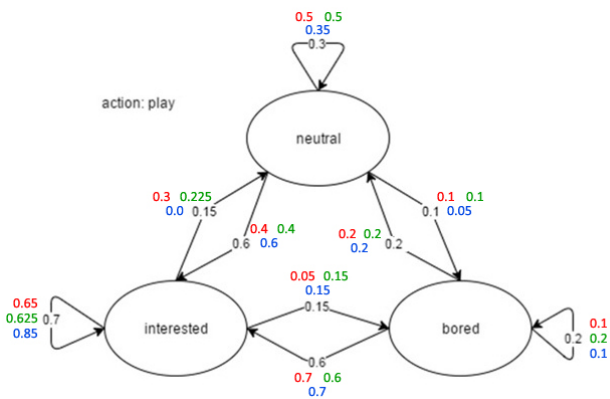


Figure 7: Depiction of the changes to the initial belief values after three sessions of interaction with the three participants, extracted from [28].

CONCLUSION

The goal of our study was to attempt to implement a new architecture for the iCub which would equip it with abilities for autonomous cognitive interaction with humans. Through detecting and tracking the affective state of the users, while simultaneously developing and maintaining its own internal evaluation system, iCub will have an effective system for evaluating the desirability of each possible action at a time, by calculating which action would make it happier with the highest probability.

A planned expansion for the perception module is to also look into utilizing information for the gaze of the person as a measurement of the level of engagement they're expressing, which if successful could be a step towards two-dimensional evaluation of the user state.

In addition to expanding the perception component, we also need to develop novel groups of differing behaviours for the iCub to perform with the participants in the experiments. We would like to monitor the emerging behaviour of iCub with each of the users, and note how the belief networks modify specifically for each one of them. Our hope for this architecture is that it will provide for a more natural, long-term, adaptive interaction between humans and robots.

REFERENCES

1. Ackovska, N. Taxonomy of learning agents, 2010.
2. Adam, C., Johal, W., Pellier, D., Fiorino, H., and Pesty, S. Social human-robot interaction: A new cognitive and affective interaction-oriented architecture. In *International Conference on Social Robotics*, Springer (2016), 253–263.
3. Baltrušaitis, T., Robinson, P., and Morency, L.-P. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, IEEE (2016), 1–10.
4. Becker-Asano, C. *WASABI: Affect simulation for agents with believable interactivity*, vol. 319. IOS Press, 2008.
5. Beer, J., Fisk, A. D., and Rogers, W. A. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction* 3, 2 (2014), 74.
6. Breazeal, C. Toward sociable robots. *Robotics and autonomous systems* 42, 3 (2003), 167–175.
7. Breazeal, C., and Scassellati, B. How to build robots that make friends and influence people. In *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*, vol. 2, IEEE (1999), 858–863.
8. Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 international conference on Multimodal interfaces*, ACM (2009), 119–126.

9. Damasio, A. R., and Macaluso, F. *L'errore di Cartesio: emozione, ragione e cervello umano*. Adelphi, 1995.
10. De la Torre, F., and Cohn, J. F. Facial expression analysis. In *Visual analysis of humans*. Springer, 2011, 377–409.
11. Dias, J., Mascarenhas, S., and Paiva, A. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*. Springer, 2014, 44–56.
12. Ekman, P., Friesen, W. V., and Hager, J. C. Facial action coding system (facs). *A technique for the measurement of facial action*. Consulting, Palo Alto 22 (1978).
13. Han, K., Yu, D., and Tashev, I. Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech* (2014), 223–227.
14. Jimenez, F., Yoshikawa, T., Furuhashi, T., and Kanoh, M. An emotional expression model for educational-support robots. *Journal of Artificial Intelligence and Soft Computing Research* 5, 1 (2015), 51–57.
15. Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., and Alami, R. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* (2016).
16. Lowe, R., and Kiryazov, K. Utilizing emotions in autonomous robots: An enactive approach. In *Emotion modeling*. Springer, 2014, 76–98.
17. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, IEEE (2010), 94–101.
18. McColl, D., Hong, A., Hatakeyama, N., Nejat, G., and Benhabib, B. A survey of autonomous human affect detection methods for social robots engaged in natural hri. *Journal of Intelligent & Robotic Systems* 82, 1 (2016), 101–133.
19. McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J., and Picard, R. 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* (2013), 881–888.
20. Michalowski, M. P., Sabanovic, S., and Simmons, R. A spatial model of engagement for a social robot. In *Advanced Motion Control, 2006. 9th IEEE International Workshop on*, IEEE (2006), 762–767.
21. Palinko, O., Rea, F., Sandini, G., and Sciutti, A. Eye gaze tracking for a humanoid robot. In *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, IEEE (2015), 318–324.
22. Palinko, O., Rea, F., Sandini, G., and Sciutti, A. Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, IEEE (2016), 5048–5054.
23. Pan, Y., Shen, P., and Shen, L. Speech emotion recognition using support vector machine. *International Journal of Smart Home* 6, 2 (2012), 101–108.
24. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
25. Sariyanidi, E., Gunes, H., and Cavallaro, A. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 6 (2015), 1113–1133.
26. Schwarz, J., Marais, C. C., Leyvand, T., Hudson, S. E., and Mankoff, J. 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*, ACM (2014), 3443–3452.
27. Sloman, A. Varieties of affect and the cogaff architecture schema. In *Proceedings of the AISB01 symposium on emotions, cognition, and affective computing. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, vol. 58 (2001).
28. Tanevska, A., Rea, F., Sandini, G., and Sciutti, A. Can emotions enhance the robot’s cognitive abilities: a study in autonomous hri with an emotional robot. In *Proceedings of AISB Convention 2017: Symposium on Computational Modelling of emotion* (2017).
29. Vaufreydaz, D., Johal, W., and Combe, C. Starting engagement detection towards a companion robot using multimodal features. *Robotics and Autonomous Systems* 75 (2016), 4–16.
30. Vernon, D. *Artificial cognitive systems: A primer*. MIT Press, 2014.