International Journal of Social Robotics
https://doi.org/10.1007/s12369-020-00661-w

# Facilitating the Child–Robot Interaction by Endowing the Robot with the Capability of Understanding the Child Engagement: The Case of Mio Amico Robot

**Chiara Filippini[1] · Edoardo Spadolini[2] · Daniela Cardone[1] · Domenico Bianchi[3] · Maurizio Preziuso[3] · Christian Sciarretta[3] · Valentina del Cimmuto[3] · Davide Lisciani[4] · Arcangelo Merla[1,2]**

Accepted: 13 May 2020


**Abstract**
Social Robots (SRs) are substantially becoming part of modern society, given their frequent use in many areas of application including education, communication, assistance, and entertainment. The main challenge in human–robot interaction is in achieving human-like and affective interaction between the two groups. This study is aimed at endowing SRs with the capability of assessing the emotional state of the interlocutor, by analyzing his/her psychophysiological signals. The methodology is focused on remote evaluations of the subject's peripheral neuro-vegetative activity by means of thermal infrared imaging. The approach was developed and tested for a particularly challenging use case: the interaction between children and a commercial educational robot, Mio Amico Robot, produced by ListianiGiochi©. The emotional state classified from the thermal signal analysis was compared to the emotional state recognized by a facial action coding system. The proposed approach was reliable and accurate and favored a personalized and improved interaction of children with SRs.

**Keywords** Human-machine interaction · Infrared imaging · Social robot · Emotions · Computational psychophysiology

## 1 Introduction

The use of social robots (SRs) has significantly increased in the last few decades, and has found applications in many fields, from healthcare for the elderly population [1, 2] to interactions with young people [3, 4].

An increasing number of studies confirmed the promise of SRs in educating and tutoring children [5, 6]. SRs are designed to interact with children in a natural, interpersonal manner, often with specific social-emotional goals in mind [7]. One example of SR with social-emotional purpose is CosmoBot, designed to interact with children with and without disabilities. CosmoBot motivates children to develop new skills more quickly than through a traditional therapy session by imitating human joint movement in its shoulders, arms, hands, and head [8]. Another example is the Kaspar robot, developed for autistic children. It represents an attempt to help in improving their skills in social interaction [9]. More recently, KineTron, a robot that coaches children with cerebral palsy to encourage their motor training [10].

The introduction of SRs into any educational practice involves several technical challenges such as an SR capable of fluent and contingent interaction. To accomplish this, the seamless integration of a range of processes in artificial intelligence and robotics are crucial. Recently, a considerable amount of effort has been dedicated towards SRs capable of achieving natural interactions. However, SRs still have significant limitations. For instance, until recently, robots had mainly followed a pre-set script, which often did not follow the normal rules of interaction, causing awkward and unnatural conversations. Moreover, the robot needs a sufficiently accurate interpretation of the social environment to respond appropriately [11].

Natural interactions require the recognition and understanding of human emotions by the robotic system, so that it is able to appropriately respond [12]. Leyzberg et al. [13,

✉ Chiara Filippini
chiara.filippini@unich.it

[1] Department of Neurosciences, Imaging and Clinical Sciences, University G. d'Annunzio of Chieti-Pescara, Chieti, Italy

[2] Next2U s.r.l., Pescara, Italy

[3] Ud'Anet s.r.l., Torrevecchia Teatina, Italy

[4] Lisciani Giochi s.r.l., Teramo, Italy

Published online: 12 June 2020

14] showed that robots that personalize content delivery, based on user satisfaction during an interaction, increase cognitive learning gains.

A conventional approach for emotion recognition in human–robot interaction (HRI) is based on facial expression analysis [15–17]. The most widely used method in this field is the facial action coding system (FACS) [18]. FACS is based on facial anatomy and it operates under the assumption that emotions activate micro-expressions, resulting in subtle changes in facial muscles activity [19]. For this purpose, FACS defines individual components of muscle movement, i.e. the Action Units (AU) [20]. AUs are reliably associated with distinct emotions [19], basing on the six universal facial expressions (happiness, anger, disgust, sadness, fear, surprise). FACS is recognized as the most common method for emotion recognition in HRI [21–23], thanks to its universality in the interpretations of the emotions.

The main contribution of the present study was to develop an alternative method for emotion recognition, relying on new metrics and evaluations obtained from the interaction between a real robot and a child. This would have ensured a substantial improvement in the field of socially contingent interactions.

The novel approach, here described, used functional Infrared Imaging (fIRI), which allows to estimate the psychophysiological state of an individual by contact-free recordings of cutaneous temperature [12].

fIRI has already been adopted in a variety of studies involving human emotions such as startle response, empathy, guilt, embarrassment, stress, fear, anxiety, and joy [24–26]. In fact, measuring facial cutaneous temperature and its topographic distribution provides insights about the person's autonomic activity. This is a result of the autonomic nervous system's (ANS) role in the achievement of the human body's thermal homeostasis and in the regulation of physiological responses to emotional stimuli [27]. The ANS has two interacting systems: the parasympathetic and sympathetic systems, which usually perform counter-balancing actions on the temperature regulation [28]. The general response to psychophysical stress or adverse conditions is the activation of the sympathetic nervous system, leading to peripheral vasoconstriction and, therefore, to a decrease in local temperature [29, 30]. In contrast, during rest or pro-social activity, the parasympathetic nervous system predominates, leading to vascular relaxation accompanied by a gradual temperature rise [31].

Concerning the literature on the evaluation of the psychophysiological state of the children using thermal infrared imaging, some works have been published in the last decades. They mainly concerned guilt [25], the reaction to a stressful situation [27, 32, 33], up to the impaired emotional regulation in children suffering from Moebius syndrome [34].

Compared to conventional techniques, fIRI relies on involuntary biological signal changes for emotion detection with the additional advantage of accessing the child's psychophysiological state, in a contact-less fashion. Furthermore, emotion detection, based on the continuous monitoring of the physiological signals, offers a solution that is free from the artifact of social masking.

To the state of the art, the only study using fIRI during child-robot interaction is represented by the Robot AVatar thermal Enhanced (RAVE) prototype project [35]. The project involved a robot which engaged babies' interest and identified when babies were "ready to learn" by classifying their facial thermal responses [36].

The solution, proposed here, relied on the findings reported in [35] and it was designed for applications where the cooperation with the subject was not guaranteed. It consisted of a Computational Psychophysiological Module (CPM), able to assess the temperature modulations in a specific facial Region Of Interest (ROI), i.e. the nose tip, and discriminate, in real-time, three macro-levels of their emotional engagement. The three macro-levels were: positive engagement, macroscopically associated with increasing temperature of the child's nose tip; neutral engagement, associated with a constant trend of the nose tip temperature; negative engagement, associated with a decreasing trend of the nose-tip temperature, thus suggesting an increasing level of stress.

The real-time processing for computational psychophysiology by means of thermal IR imaging in the realistic scenario has been already demonstrated [37–39] by employing high-end thermal infrared cameras. In this perspective, the present study aimed to develop a viable solution for social robots integrating consumer market technology and low-cost Original Equipment Manufacturer (OEM) based components. Proper computational methods were developed for the goal of providing a commercial low-cost educational robot, adapting its behavior to the engagement level of the interacting child.

## 2 Materials and Methods

### 2.1 Participants

The experimental session involved 31 children, aged from 4 to 5 years old, including 4 children with social interaction difficulties. The study was conducted in the primary school "G. Rocchetti" of Torrevecchia Teatina (CH)—Italy, in the Italian language.

Before the start of the experimental trials, the parents were widely informed about the purpose and protocol of the study and they signed an informed consent form.

## 2.2 Materials and Data Acquisition

The SR employed in this study was Mio Amico Robot produced by Liscianigiochi© (https://www.liscianigroup.com/mio-amico-robot-interattivo/scheda/24435).

The robot is equipped with voice recognition, activated from up to 5 meters away, and with the Text To Speech (TTS) program, allowing semantic analysis and speech synthesis. These systems allow the robot to interact with the child. When the robot asks a question, the "listening" mode is activated immediately, waiting for the child's answer. The robot, after listening to the command, performs the required action. Moreover, it can move in all directions (forward, backward, right, left), turn on itself and move its head.

The most important hardware devices of Mio Amico are listed below (Table 1).

For the purpose of this study, Mio Amico was equipped also with a radiometric OEM thermal camera, FLIR Lepton 3.5® (long-wave infrared (LWIR) sensor, uncooled Vanadium Oxide (Vox) microbolometer, thermal sensitivity < 50 mK, frame rate < 9 Hz, 160 × 120 pixel matrix). It is a micro-thermal camera with the dimensions of 11.8 × 12.7 × 7.22 mm. It was embedded in the head of Mio Amico, next to the robot webcam, to ensure vertical alignment and equal orientation of both imaging devices.

The SR processing unit and the thermal camera communicated via USB, using the GetLab PureThermal2 I/O module. The module handled powering and low-level communication with the Lepton camera. It allowed control and data capturing from the processing unit side using the standard USB Video Class (UVC) protocol. The exchanged data was the full radiometric 14-bit raw data from the sensor (including all relevant telemetry data).

Both the visible and thermal data were processed and stored on the ODROID XU4 processing unit of the SR (Fig. 1).

## 2.3 Procedure

According to the International Academy of Thermology (IACT) guidelines, prior to the measurement session, children acclimated for about 15 min inside the experimental room. The room was kept at steady temperature and humidity (23 ± 1 °C; 50–60% relative humidity) and without any direct air ventilation on the participants [26].

The experimental protocol, summarized in Fig. 2, was composed of an "event-related" paradigm. Specifically, the events consisted of the following two types of actions:

1. Robot telling a fairy tale;
2. Robot singing a song.

The choice of one or the other action depended on the will of the child.

**Table 1** List of Mio Amico hardware devices and functionalities

| Hardware device | Functionality |
|---|---|
| ELP 5MP | Webcam |
| Tactile sensor | Located behind the robot head. It allows the recognition of caresses and it is used to turn the robot on |
| Distance sensor | It allows the robot to avoid obstacles when it is moving |
| Motors | Movement of wheels, arms, and neck |
| Inertial measurement unit (IMU) | Ultrasonic and temperature sensors |
| Microphone and speakers | Oral interaction with the child |
| Led matrix for the face | Led matrix smiley-shaped, used as robot face |
| Pack batteries | Power supply |

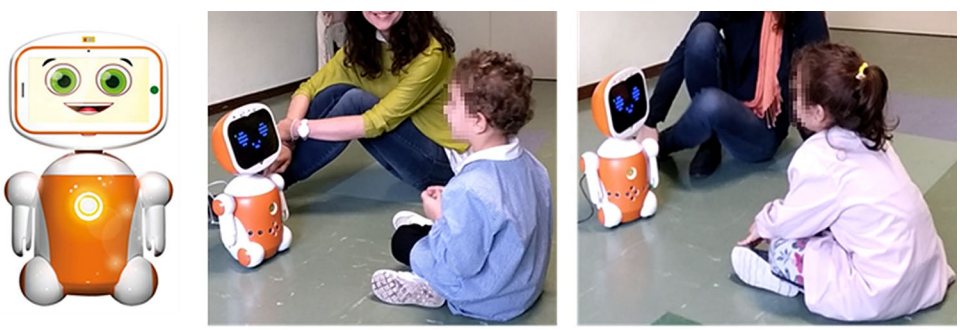

**Fig. 1** **a** Mio Amico Robot; **b**, **c** two examples of child and robot interaction under the supervision of an adult

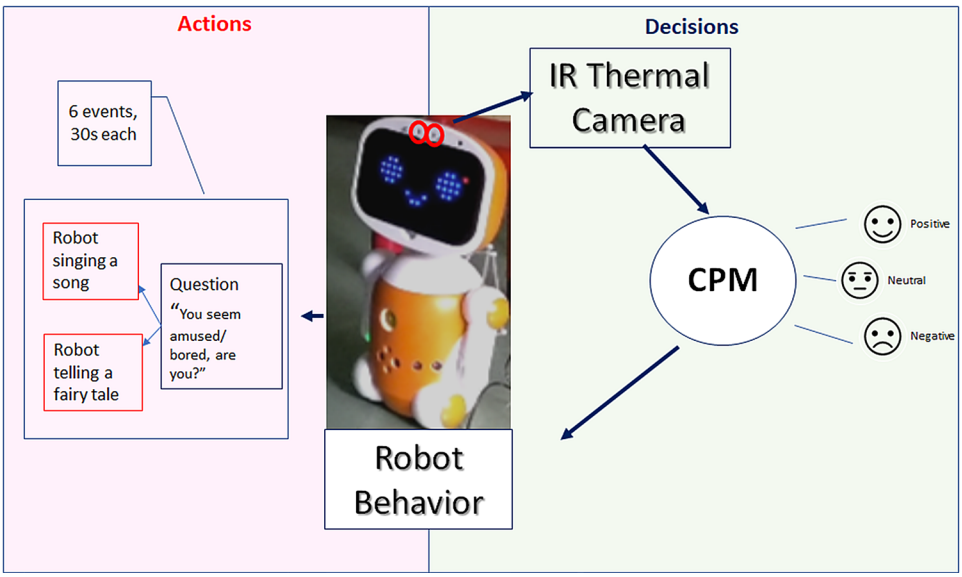


**Fig. 2** Procedure overview. The behavior of the robot depended on the CPM outcome, which takes its input from the thermal camera. The robot's behavior consisted of two types of actions: robot telling a fairy tale or robot singing a song. At the end of each event, the robot asked the child a question. Depending on the CPM outcome the robot's question could have been "You seem amused, right?" or "You seem bored, right?". Based on the child's response, the next action was selected

The experimental protocol was structured in two phases. In the first phase (*Familiarization*), the SR introduced itself and asked the child the following sequence of questions: "What's your name?"; "How old are you?";"How are you today?";"Do you want to play with me?". At the end of each question, the robot waited for the child's answer before asking another question.

The second phase (*Interaction*) consisted in the presentation of the actions. The robot started telling a fairy tale, lasting about 30 s. At the end of the action, the robot queried the CPM module:

- If the prevailing status of the child (identified by the CPM) during the robot action was positive then the SR asked: "You seem amused, right?"
- If the prevailing state was negative, then the SR asked: "You seem bored, right?"

Following the child's response, the type of the next action was selected:

- If the child confirmed that he/she was amused, the action remained the same;
- If the child said he/she was bored, the action changed and, in this case, a song was sung.

At the second occurrence of the same action, i.e. after listening to two consecutive fables or two songs, if the child still seemed amused the robot asked:

"Do you want to hear another fairy tale, or do you want me to sing you a song?"

Each experimental test consisted of 6 events. The time between the end of the event and the beginning of the robot question was about 1.5 s (i.e. waiting time + random delay).

At the end of the sixth event, as a conclusion of the session and as confirmation of the child's general satisfaction with the SR, it said:

"Well, for today our time together is over, did you enjoy playing with me? … Thank you, my friend".

### 2.4 Behavioral Data Analysis

The analysis of the expressive component of emotion starts from the assumption that the different emotions are correlated with specific configurations of the face. The analysis of facial expressions was performed through the software FaceReader 7 (developed by VicarVision and Noldus Information Technology, https://www.noldus.com/human-behavior-research/products/facereader). FaceReader 7 is a software for automatic recognition of FACS. It has already been used in literature and validated in various research studies [40].

By using FaceReader 7, the steps in recognizing emotions are represented by the identification of the face, then by its modeling and lastly by the expressive classification of the emotions (Fig. 3).

For the specific aim of this study, the valence and arousal indices, identified by the FaceReader 7 software, were taken into account. Valence index, on the one hand, indicates whether the subject's current emotional state is positive or negative. Joy is the only emotion considered entirely positive, whereas sadness, anger, fear, and disgust are considered negative. Surprise, instead, can be considered both

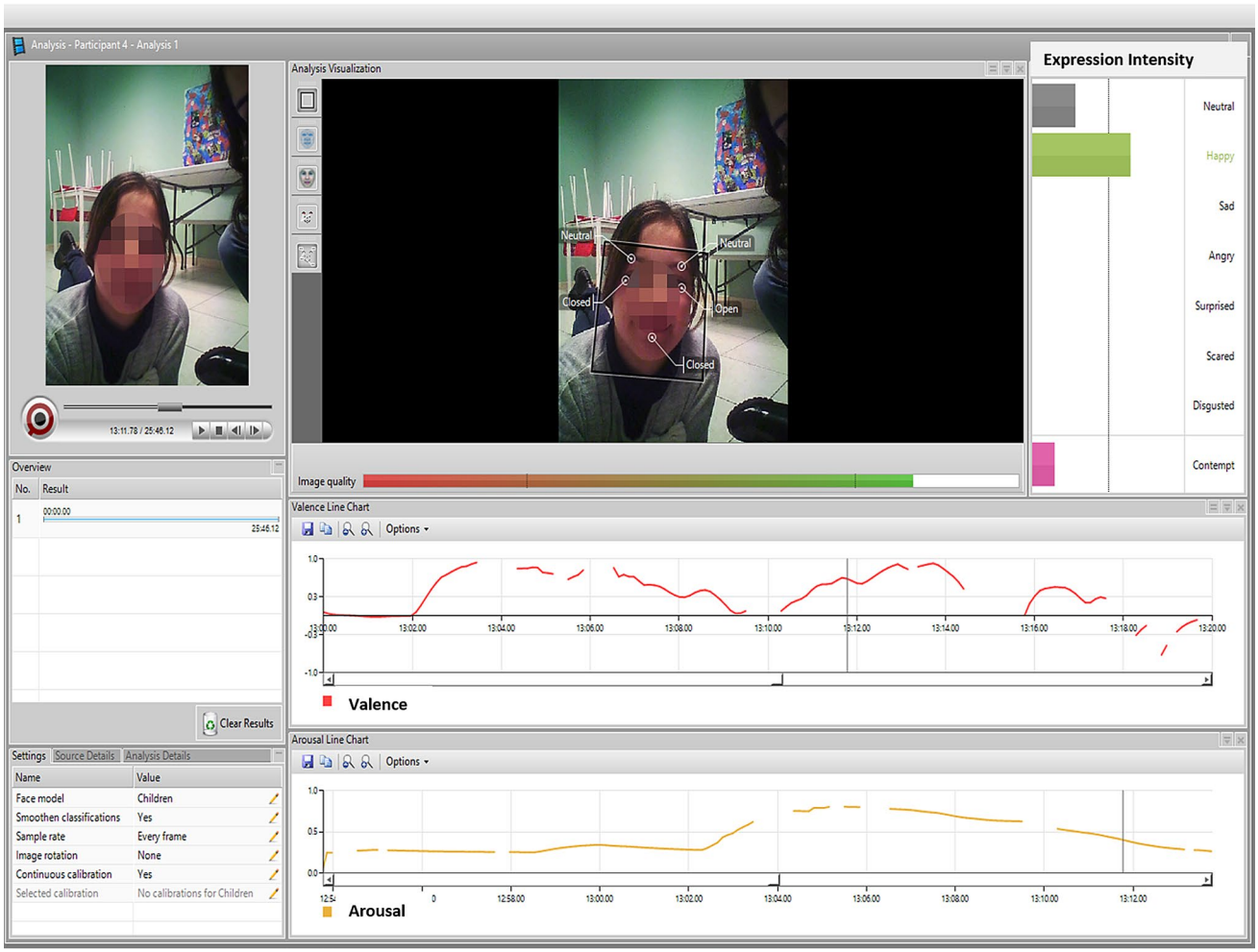


**Fig. 3** Example of analysis of a visible video by means of FaceReader 7

positive and negative. Arousal index, on the other hand, indicates whether the subject is responsive or not, at that given moment and for that given stimulus, and how active he/she is. The product of these two behavioral indices (VA) was integrated into a decision-making model, which allowed to provide outgoing discrimination between three states (positive, negative and neutral). This model was based on the widely used "Circumplex Model of Affect" [41], which has valence and arousal indices as dependent variables. For instance, according to the Circumplex Model, "excited" is an affective state with high arousal and positive valence (high positive VA), while "angry" describes a high arousal state and negative valence (high negative VA). Therefore, for the purpose of this study, the three-level of emotional engagement were identified as follows:

- Positive engagement: VA's positive values (i.e., $VA > 0$);
- Neutral engagement: a low negative value of VA (i.e., $-0.1 < VA < 0$), since for the requirement of the study, neutral engagement was considered as a sort of disengagement.
- Negative engagement: VA's values lower than $-0.1$ (i.e., $VA < 0.1$).

The principal limits of this behavioural analysis approach were: (1) adequate image contrast; (2) partial or total occlusion of the face from the scene.

## 2.5 Computational Psychophysiology Module (CPM)

The Computational Psychophysiology Module (CPM) is a tool that assesses temperature modulations within facial ROIs. Based on such modulations, the tool was able to classify, in real-time, three macro-levels of the child's emotional engagement. The CPM used visible and infrared imaging sensors, both embedded in the robot head. The visible images were employed for face detection and landmarks localization through state-of-the-art computer

vision algorithms [42, 43]. Landmarks location in the visible images were then transformed in thermal images coordinates, relying on an optical calibration approach. Landmark coordinates were used to identify and track facial ROIs in the thermal images. ROIs' average temperatures over time were estimated, processed and utilized for classification. The real-time classification was used by the robot's main program to identify the psychophysiological state. For the sake of clarity, the only nose tip ROI was considered in this study. Further explanations are available in 2.5.3 section.

In the next sections, visible and thermal data analysis and the classification of thermal responses will be discussed separately.

#### 2.5.1 CPM: Visible and Thermal Data Co-registration

A preliminary step, in the described work, consisted in an optical calibration procedure between the two imaging systems, i.e. visible and thermal devices.

First, it was necessary to calculate the intrinsic and extrinsic parameters. The former ones are specific for each sensor, whereas the latter is related to the specific geometrical transformation between the two imaging coordinates systems.

The typical procedure for estimating these parameters relies on the location of a known configuration of points, seen from different angles and simultaneously acquired by the two imaging devices. Typically, the corners of a black and white checkerboard are localized. Knowing the dimensions of the checkerboard's squares, the parameters are estimated as to allow an optimal correspondence between the pixel coordinates in the image and the coordinates in real units.

For the purpose of this study, a special checkerboard, whose details were clearly detectable by both the visible-spectrum and the thermal camera was designed.

The developed solution consisted of an adhesive decal made of black vinyl with a cutting plotter, stuck to an aluminum plate (Fig. 4a). The details of the checkerboard were clearly detectable in the thermal images, given the different emissivity values of aluminum and plastics ($\varepsilon_{\text{aluminium}} = 0.090$; $\varepsilon_{\text{plastics}} = 0.950$ [44]) (Fig. 4b).

The optical calibration process used already developed procedures, implemented in OpenCV [45]. An example of the application of the distortion correction obtained through the calibration process is showed in Fig. 4c.

Generally, optical calibration is valid at a fixed distance. To overcome this constraint, the distance of the face from the camera was estimated over time, basing on an anatomical model. The results appeared to be adequately precise for subsequent analyses, in the range of distances involved.

The calibration process has to be done only once before the whole experimental session.

#### 2.5.2 CPM: Thermal Data Extraction and Analysis

For the extraction of physiological signals, an automatic data processing pipeline was implemented. Signal processing techniques were chosen on the basis of their efficiency in terms of computational load, to allow acceptable performance for real-time processing.

The first phase relied on the visible image to detect the child's face and locate specific facial landmarks, corresponding to the eyebrows, the edges of the eyes, the nose tip, the chin. For the face localization, an object detector based on the histogram of oriented gradients (HOG) [42] was used. The extraction of landmarks was based, instead, on a regression tree ensemble algorithm. From the coordinates extracted in the visible frame, the corresponding positions in the thermal image were obtained, thanks to a previous procedure of optical calibration (ref. Section 2.5.1).

Subsequently, thermal signals were extracted from the nose tip region. The extracted raw signal was not immediately suitable for the analysis, due to both the intrinsic noise, given the low resolution of the thermal sensor, and the possible temporary absences of data. It was possible to compensate for both problems, without losing relevant information for psychophysiological purposes. A constant sampling frequency was guaranteed, interpolating linearly in each missing signal segment, lasting less than a second. A 3-sample boxcar FIR filter was applied to the reconstituted signal, to compensate for the low signal to noise ratio (SNR) of the thermal data.

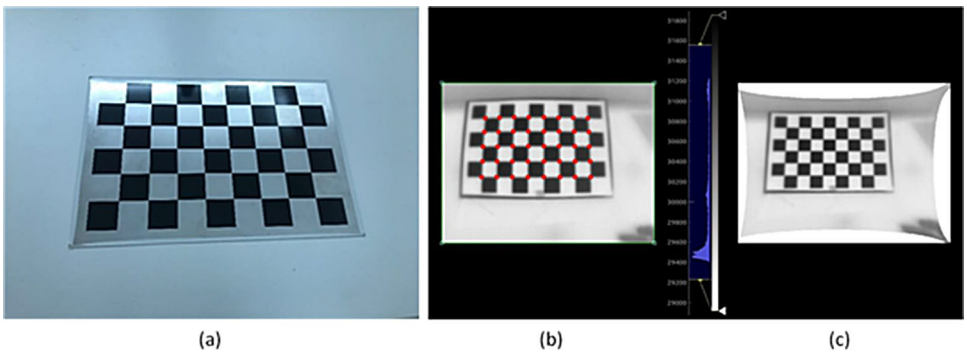


**Fig. 4** Checkerboard used for calibration: **a** visible image; **b** thermal image with detected edges; **c** thermal image after the distortion correction

The processed signals were used as indicators of activation of the sympathetic/parasympathetic system.

An example of the interface, developed for the extraction of the thermal signals, is shown in Fig. 5.

### 2.5.3 CPM: Classification of the Psychophysiological Response

To calculate information about the child's affective state, the processed thermal signal of the nose tip region was used. The classification task was performed using a data-driven approach, guided by the behavioral analysis outcome. In detail, the dynamic of the nose tip thermal signal was classified based on the VA index (ref. Section 2.4). The only nose tip region was considered as the salient ROI because of its strict neurovascular relationship with adrenergic activity, associated with expression of emotional states [30].

Classification problems are well suited for machine learning approaches. Although many neural network (NN) models have been developed, the multi-layer perceptron (MLP) feed-forward neural network is still extensively used [46]. A NN consists of units (neurons), arranged in layers, which convert an input vector into some output [47]. NN is defined feed-forward since a unit feeds its output to all the units on the next layer. A three-layer structure, widely validated in literature, was employed in the present study, given its capability to solve most classification problems. The three layers include one input layer, one hidden layer, and one output layer. Each layer is composed of several units, all connected with each other, except for the units in the same layer. The input layer, the hidden layer, and the output layer

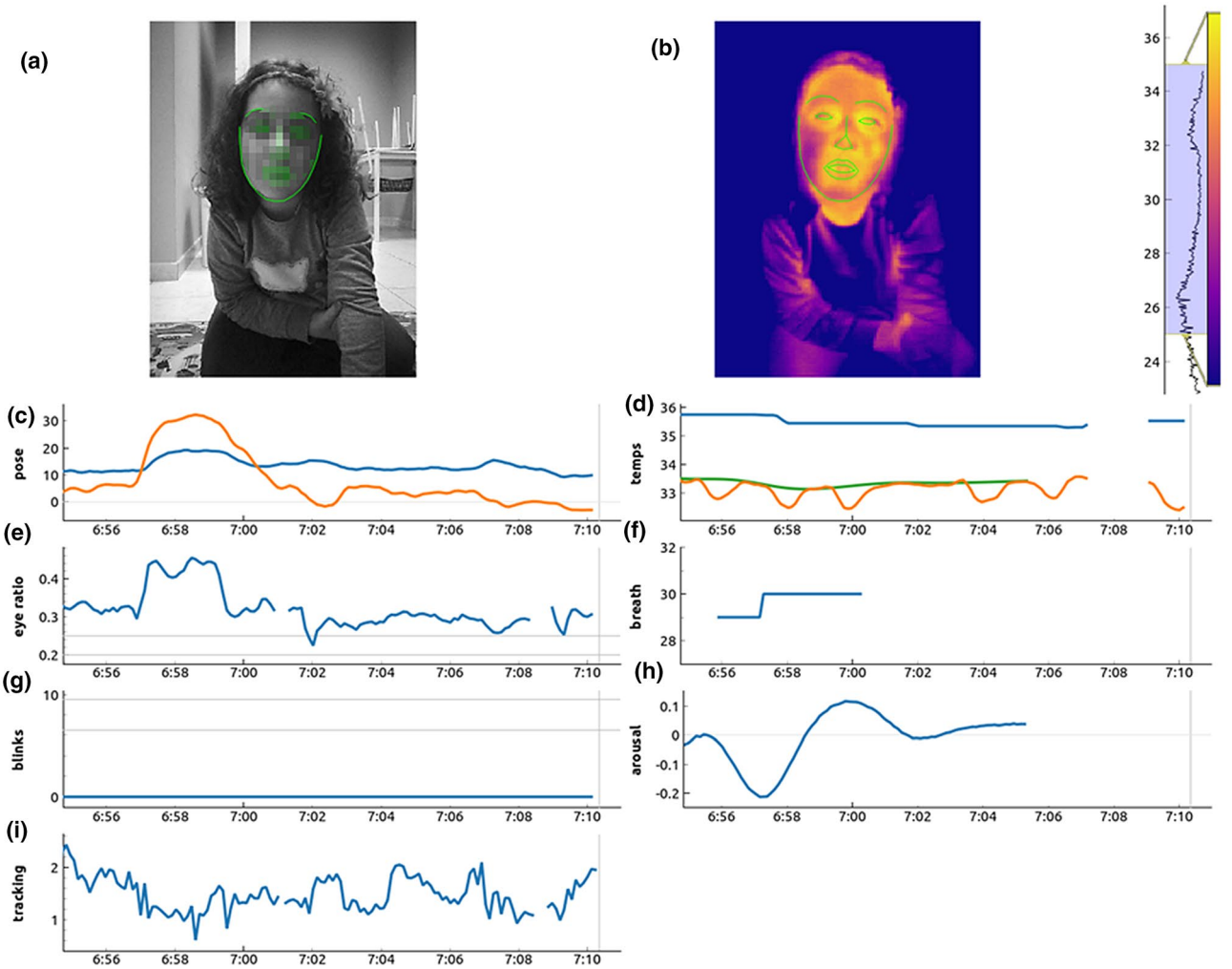


**Fig. 5** Example of analysis of thermal signals. The coordinates of the landmarks detected in the visible image (**a**) are reported on the corresponding thermal image (**b**). The data processed by the CPM are: pose (**c**), ROIs temperature (**d**), eye ratio (**e**), breath (**f**), blinking (**g**), arousal (**h**), tracking performance (**i**)

are used for data input, data processing, and data output, respectively. Data processing or training involves adjusting the parameters, or the weights and biases, of the model to minimize the classification error [48]. The desired output was identified as the emotional state predicted by the VA index (ref. Section 2.4). Input data, instead, consisted of the first-time derivative of the signal, which provides information about how quickly the temperature over time.

In detail, accounting for the delay in temperature response, an interval of 5 s was considered sufficient to discriminate changes in emotional state. Therefore, the input data were set as the average value of the temperature's first derivative in 5 s. The desired output was defined as the emotional state, detected by the VA index, prevalently occurring during the same period of time.

MLP algorithm was run in SPSS 25.0 (SPSS Inc., Chicago, IL, USA). The dataset has been partitioned into training and testing samples, 70%, and 30% respectively, randomly assigned. The used activation function on the output layer was the softmax, whereas the gradient descendent was chosen as optimization algorithm.

## 3 Results

Concerning the extraction and analysis tools of CPM, the performance of the described procedure was very high, as only a few samples per video were lost (on average, 82.75% of the thermal data was correctly tracked and available for successive analysis). Besides, through the described algorithms and the employed hardware (single-board Odroid XU4), a high speed of extraction and processing of the signals was guaranteed (~20 frames per second).

Concerning the classification process (described in Sect. 2.5.3), the results of the MLP analysis are reported in Fig. 6, showing the Receiver Operating Characteristic (ROC) curve for each category. Each curve treats each category as the positive state versus the aggregate of all the other categories. By analyzing the coordinate points of each ROC curve, the classifier thresholds were established, to accomplish a compromise between sensitivity and specificity. These thresholds were applied to the average value of the first-time temperature derivative in 5 s, to switch from a positive, neutral or negative emotional state.

Table 2 shows the specific threshold levels (points marked with an asterisk in Fig. 6) applied to the average value of temperature's first-time derivative (avg_TD).

The thresholds relative to the neutral state derived from the positive and negative threshold levels; therefore, the value of avg_TD between − 0.004 and 0.007 fell in neutral emotional state.

The overall MLP accuracy is described through the confusion matrix reported in Table 3.

An overall level of accuracy of 71% was reached, while a precision level of 69%, 60%, and 77% is ensured for positive, neutral and negative emotional states respectively.

Figure 7 shows the emotional state of a random subject obtained by means of FaceReader 7 and the emotional state of the same subject, obtained as a result of MLP processing.

A representative example of CPM output and performance in positive, neutral or negative emotional state recognition is shown in Fig. 8, for one randomly chosen subject.

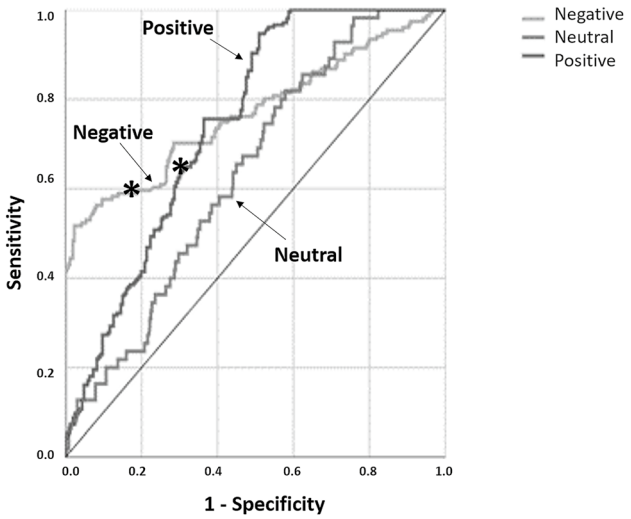


**Fig. 6** ROC curves representing the MLP outcome performed through SPSS. The ROC curve illustrates the true positive rate against the false-positive rate at various threshold settings. The points marked with an asterisk represent the threshold cut-off that maximizes (sensitivity + specificity) for each curve. The values of each cut-off are reported in Table 2. The 45° line represents the no-discrimination line

**Table 2** Thresholds obtained as a cut-off point from ROC curve analysis

| Class | Thresholds | Specificity | Sensitivity |
|---|---|---|---|
| Positive | avg_TD > 0.007 | 0.73 | 0.67 |
| Negative | avg_TD < −0.004 | 0.82 | 0.60 |

**Table 3** Confusion matrix visualizing the performance on the 17-person validation set. The overall accuracy of the CPM module prediction was 71%

| Ground truth | CPM Prediction | | | |
|---|---|---|---|---|
| | Positive | Neutral | Negative | |
| Positive | 51 | 2 | 5 | Accuracy 71% |
| Neutral | 12 | 15 | 9 | |
| Negative | 10 | 8 | 48 | |
| Precision | 69% | 60% | 77% | |

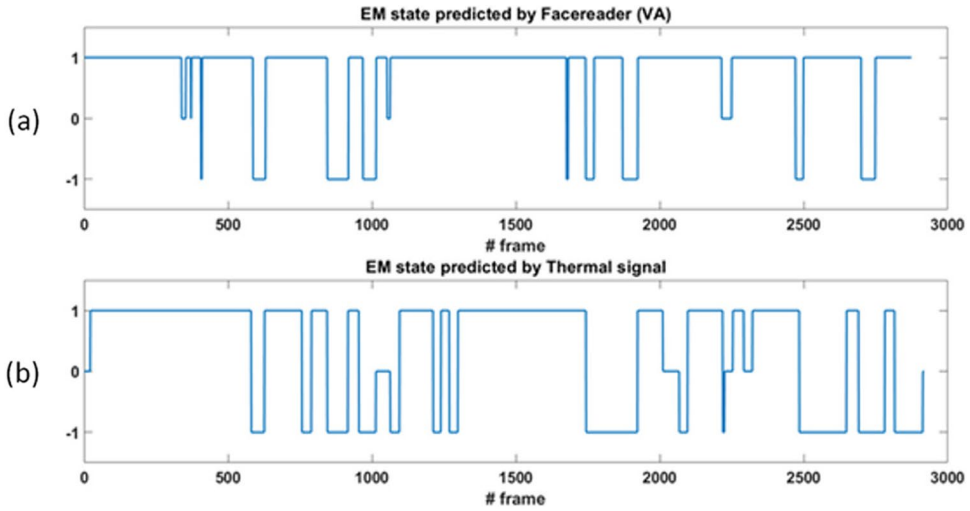


**Fig. 7** The emotional state of a random subject obtained by means of FaceReader 7 (**a**) and the emotional state of the same subject, obtained from MLP processing (**b**)

The filtered nose tip thermal signal and the inferred thermal emotional state of the subject are also reported in Fig. 8.

For the sake of clarity, only 17 videos, out of a total of 31, were processed and were part of the validation analysis of the procedure. The main reasons for the video to be excluded were:

- problems in face detection, i.e. the child's face was not recognized by the face detector module, because of the low quality of the visible video;
- breathing artifacts, i.e. the thermal signal of the nose tip was heavily corrupted by the breathing signal. The two contributes of signals could not be separable, given also the low thermal resolution of the employed thermal camera;
- excessive movements of the robot head.

## 4 Discussion

In the last decades, SRs have been widely employed with children for many purposes, i.e. education, health, communication. However, some limitations have been observed during human–robot interaction, mainly due to non-natural cooperation between the two parts.

In the present study, a novel methodology, allowing for a suitable and natural interaction between SR and children is presented. Using the developed method, it was possible to classify, with a sufficient level of accuracy, the engagement state of the child, while interacting with an artificial agent. To achieve this goal, the thermal facial response of children, i.e. nose tip temperature signal, was monitored and categorized in real-time during an experimental session with Mio Amico Robot. The classification was carried out using a data-driven approach, relying on FACS, which is recognized as gold standard in emotion recognition in HRI [21–23]. By comparing the emotional state classified from thermal signal analysis with the emotional state recognized by Facereader 7 (VA index), the engagement level of the child was assessed.

The advantage of using fIRI compared to FACS consists of the availability of recognizing the psychophysiological state of the child, relying on involuntary biological signals measurement. Thus, it allows to avoid the artifact of social masking, making it suitable also for children who lack the ability to express emotions [34]. Moreover, basing the evaluation of the emotional state only on the anatomic information (e.g., relying only on the visible imaging) rather than the functionality of a specific facial area, could be misleading. Besides, concerning the emotional state identification, the social context of use is crucial and can constitute a bias in the canonical methods used for this purpose [49, 50]. fIRI permits to overcome these limitations, adding information also on the functionality of a facial area and allowing measurements of spontaneous and not self-regulated parameters.

Although there is a considerable interest in the use of fIRI for recognizing emotions, the relationship between the thermal signal and the two dimensions of the emotion (i.e. arousal and valence), is not well defined yet. Pavlidis et al. and Kosonogov et al. indicated that facial thermal imaging is a reliable technique for the assessment of emotional arousal [51, 52]. On the other hand, Salazar-López reported a strong correlation of thermal imaging with emotional valence [53]. In the present study, to account for both emotional dimensions, the product of the two indices was considered as representative of the child engagement. Additional data across a wider sample of emotional states would be needed to better clarify the precision of fIRI in recognizing emotions.

At the state of the art, the presented work constitutes the first step towards a more natural interaction between the artificial agent and the child, based on physiological signals. Furthermore, it was performed in real-time and in a non-invasive fashion, ensuring to maintain an ecologic condition

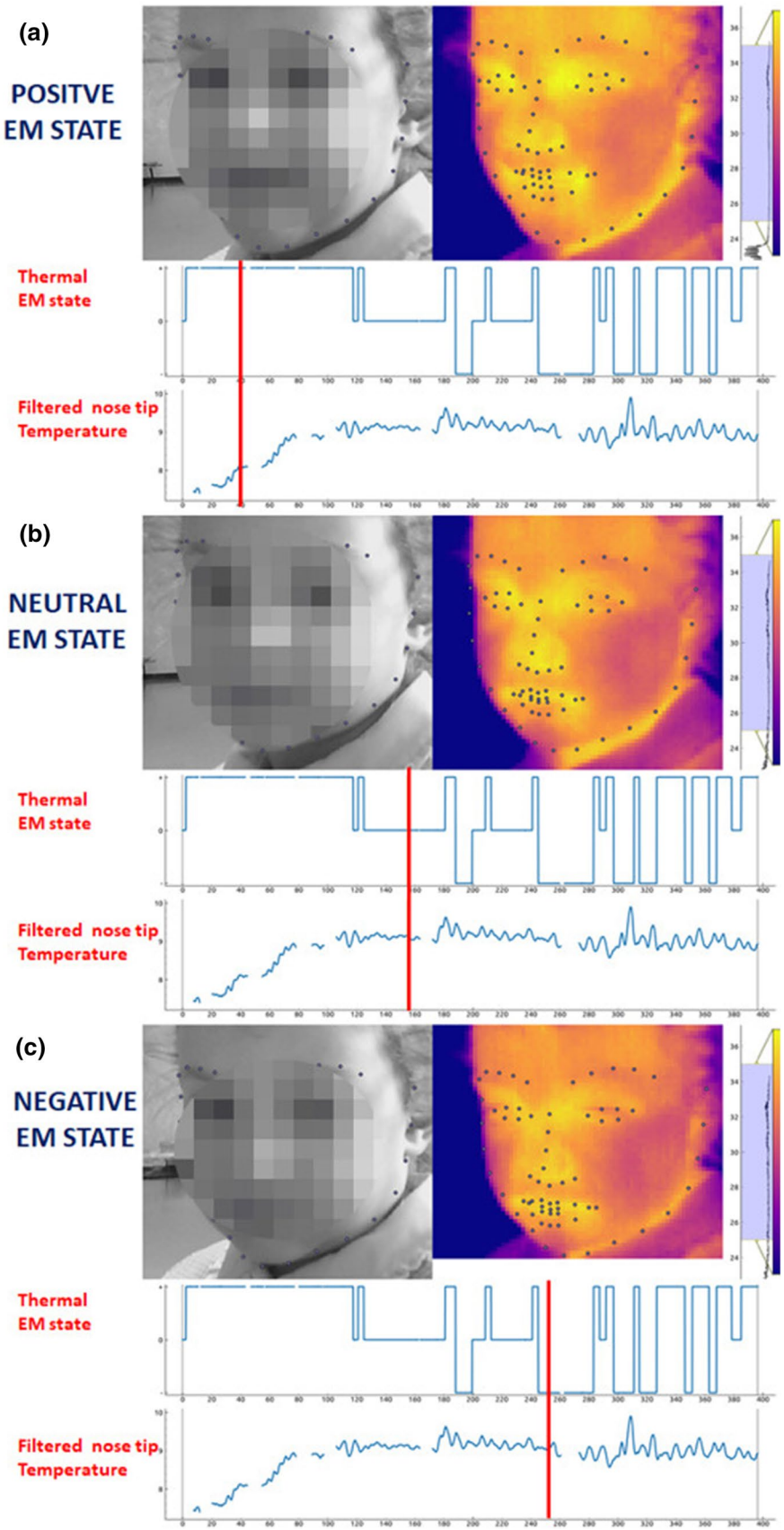


**Fig. 8** Example of the CPM performance in the detection of **a** positive **b** neutral and **c** negative emotional state. The index of the identified emotional state (Thermal EM state) and the filtered thermal signal is reported below each visible and thermal frame. The red line indicates the time point related to the visible and thermal frames

during measurements. Despite the availability of thermal signals from several ROIs (i.e. perioral area, glabella) and different additional information (i.e. head pose, blinking signals) (refer to Fig. 5), the chosen strategy relied on the contribution of the only nose-tip thermal signal. This choice accomplishes a compromise between the accuracy in prediction of emotional state and the computational load for the machinery. The nose tip region was considered because of its strict neurovascular relationship with adrenergic activity, associated with expression of emotional states [30]. Moreover, the nasal area revealed the strongest responsiveness to the facial expressions compared to other regions as forehead, cheeks and mouth areas [54].

On the other hand, however, it is important to mention that there are several limitations to consider for further improvements. First of all, it would be necessary to study a wider sample of population, to establish more accurate threshold levels for the discrimination of the emotional states. Secondly, since the real-time tracking of thermal videos relied on disposable packages for visible videos, all the limitations of the above-mentioned solutions were directly inherited by the presented method. In particular, the developed technique seems not to work properly in case of partial absence of light or partial occlusion of the subject from the scene. A further improvement would be to develop a real-time tracker, based on the only IR videos, acquired by low-resolution thermal cameras, to avoid problems due to low-light environment and to directly accede to the psychophysiology state of the human interlocutor.

## 5 Conclusion

In the present work, a novel and original method, allowing for a natural interaction between SR and children is presented. By using the developed technique and low-cost OEM thermal infrared sensors, it was possible to classify, with a sufficient level of accuracy, the child engagement, while interacting with an artificial agent. At the state of the art, the presented work constitutes the first step towards a reliable interaction between the child and the robot, based on the assessment of psychophysiological response. It was realized in real-time and in a non-invasive fashion, ensuring to maintain an ecologic condition during measurements.

**Acknowledgements** We thank the staff and pupils of the primary school “G. Rocchetti” in Torrevecchia Teatina (CH) - Italy for their support and availability during the experimental sessions.

**Funding** This study was funded by MIUR PON Project Sensing Robot—n. F/050326/01-02/X32 FONDO PER LA CRESCITA SOSTENIBILE (F.C.S.) Horizon 2020 - PON I&C 2014/2020 (D.M. 1/06/2016).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.



## References

1. Broekens J, Heerink M, Rosendal H (2009) Assistive social robots in elderly care: a review. Gerontechnology 8(2):94–103
2. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G (2013) Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas 74(1):14–20
3. Toh LPE, Causo A, Tzuo PW, Chen IM, Yeo SH (2016) A review on the use of robots in education and young children. J Educ Technol Soc 19(2):148–163
4. Boucenna S, Narzisi A, Tilmont E, Muratori F, Pioggia G, Cohen D, Chetouani M (2014) Interactive technologies for autistic children: a review. Cognit Comput 6(4):722–740
5. Mubin O, Stevens CJ, Shahid S, Al Mahmud A, Dong JJ (2013) A review of the applicability of robots in education. J Technol Educ Learn 1(209–0015):13
6. Belpaeme T, Baxter P, Read R, Wood R, Cuayáhuitl H, Kiefer B, Looije R (2013) Multimodal child-robot interaction: building social bonds. J Hum–Robot Interact 1(2):33–53
7. Breazeal C, Takanishi A, Kobayashi T (2008) Social robots that interact with people. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer, Berlin
8. Lathan C, Brisben A, Safos C (2005) CosmoBot levels the playing field for disabled children. Interactions 12(2):14–16
9. Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) KASPAR–a minimally expressive humanoid robot for human–robot interaction research. Appl Bion Biomech 6(3–4):369–397
10. Malik NA, Hanapiah FA, Rahman RAA, Yussof H (2016) Emergence of socially assistive robotics in rehabilitation for children with cerebral palsy: a review. Int J Adv Rob Syst 13(3):135
11. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. Sci Robot 3(21):eaat5954
12. Merla A (2014) Thermal expression of intersubjectivity offers new possibilities to human–machine and technologically mediated interactions. Front Psychol 5:802
13. Leyzberg D, Avrunin E, Liu J, Scassellati B (2011, March) Robots that express emotion elicit better human teaching. In: Proceedings of the 6th international conference on human–robot interaction. ACM, pp 347–354
14. Leyzberg D, Spaulding S, Scassellati B (2014, March) Personalizing robot tutors to individuals’ learning differences. In:

Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction. ACM, pp 423–430
15. Cosentino S, Randria EI, Lin JY, Pellegrini T, Sessa S, Takanishi A (2018, October) Group emotion recognition strategies for entertainment robots. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 813–818
16. Liu Z, Wu M, Cao W, Chen L, Xu J, Zhang R, Mao J (2017) A facial expression emotion recognition based human–robot` interaction system. IEEE/CAA J Autom Sin 4(4):668–676
17. Foster ME, Gaschler A, Giuliani M (2017) Automatically classifying user engagement for dynamic multi-party human–robot interaction. Int J Soc Robot 9(5):659–674
18. Menne IM, Schnellbacher C, Schwab F (2016, November) Facing emotional reactions towards a robot—an experimental study using FACS. In: International conference on social robotics. Springer, Cham, pp 372–381
19. Tracy JL, Robins RW, Schriber RA (2009) Development of a FACS-verified set of basic and self-conscious emotion expressions. Emotion 9(4):554
20. Kaiser S, Wehrle T (1992) Automated coding of facial behavior in human-computer interactions with FACS. J Nonverbal Behav 16(2):67–84
21. Kędzierski J, Muszyński R, Zoll C, Oleksy A, Frontkiewicz M (2013) EMYS—emotive head of a social robot. Int J Soc Robot 5(2):237–249
22. Breazeal C (2003) Emotion and sociable humanoid robots. Int J Hum Comput Stud 59(1–2):119–155
23. May AD, Lotfi A, Langensiepen C, Lee K, Acampora G (2017) Human emotional understanding for empathetic companion robots. In: Advances in computational intelligence systems. Springer, Cham, pp 277–285
24. Cardone D, Pinti P, Merla A (2015) Thermal infrared imaging-based computational psychophysiology for psychometrics. Computat Math Methods Med 2015:984353
25. Ioannou S, Ebisch S, Aureli T, Bafunno D, Ioannides HA, Cardone D, Merla A (2013) The autonomic signature of guilt in children: a thermal infrared imaging study. PLoS ONE 8(11):e79440
26. Cardone D, Merla A (2017) New frontiers for applications of thermal infrared imaging devices: computational psychopshysiology in the neurosciences. Sensors 17(5):1042
27. Ebisch SJ, Aureli T, Bafunno D, Cardone D, Romani GL, Merla A (2012) Mother and child in synchrony: thermal facial imprints of autonomic contagion. Biol Psychol 89(1):123–129
28. Gordan R, Gwathmey JK, Xie LH (2015) Autonomic and endocrine control of cardiovascular function. World J Cardiol 7(4):204
29. Shastri D, Merla A, Tsiamyrtzis P, Pavlidis I (2009) Imaging facial signs of neurophysiological responses. IEEE Trans Biomed Eng 56(2):477–484
30. Engert V, Merla A, Grant JA, Cardone D, Tusche A, Singer T (2014) Exploring the use of thermal infrared imaging in human stress research. PLoS ONE 9(3):e90782
31. Ziegler MG (2012) Psychological stress and the autonomic nervous system. In: Primer on the autonomic nervous system, 3rd edn, pp 291–293
32. Aureli T, Grazia A, Cardone D, Merla A (2015) Behavioral and facial thermal variations in 3-to 4-month-old infants during the Still-Face Paradigm. Front Psychol 6:1586
33. Mazzone A, Camodeca M, Cardone D, Merla A (2017) Bullying perpetration and victimization in early adolescence: physiological response to social exclusion. Int J Dev Sci 11(3–4):121–130
34. Nicolini Y, Manini B, De Stefani E, Coudé G, Cardone D, Barbot A, Bianchi B (2019) Autonomic responses to emotional stimuli in children affected by facial palsy: the case of Moebius syndrome. Neural Plast 2019:7253768
35. Scassellati B, Brawer J, Tsui K, Nasihati Gilani S, Malzkuhn M, Manini B, Traum D (2018, April) Teaching language to deaf infants with a robot and a virtual human. In: Proceedings of the 2018 CHI conference on human factors in computing systems. ACM, p 553
36. Nasihati Gilani, S., Traum, D., Merla, A., Hee, E., Walker, Z., Manini, B., … & Petitto, L. A. (2018, October). Multimodal Dialogue Management for Multiparty Interaction with Infants. In Proceedings of the 2018 on International Conference on Multimodal Interaction (pp. 5-13). ACM
37. Buddharaju P, Dowdall J, Tsiamyrtzis P, Shastri D, Pavlidis I, Frank MG (2005, June) Automatic thermal monitoring system (ATHEMOS) for deception detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 2. IEEE, pp 1179-vol
38. Dowdall J, Pavlidis IT, Tsiamyrtzis P (2007) Coalitional tracking. Comput Vis Image Underst 106(2–3):205–219
39. Merla A, Cardone D, Di Carlo L, Di Donato L, Ragnoni A, Visconti A (2011) Noninvasive system for monitoring driver's physical state. In: Proceedings of the 11th AITA advanced infrared technology and applications
40. Zaman B, Shrimpton-Smith T (2006, October) The FaceReader: measuring instant fun of use. In: Proceedings of the 4th nordic conference on human–computer interaction: changing roles. ACM, pp 457–460
41. Posner J, Russell JA, Peterson BS (2005) The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17(3):715–734
42. Dalal N, Triggs B (2005, June) Histograms of oriented gradients for human detection. In: International conference on computer vision & pattern recognition (CVPR'05), vol. 1. IEEE Computer Society, pp 886–893
43. Kazemi V, Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1867–1874
44. Chemical Rubber Company (1920) Handbook of chemistry and physics. Chemical Rubber Publishing Company
45. Bradski G, Kaehler A (2008) Learning OpenCV: computer vision with the OpenCV library. O'Reilly Media Inc, Newton
46. Zhang Z, Lyons M, Schuster M, Akamatsu S (1998, April). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: Proceedings third IEEE international conference on automatic face and gesture recognition. IEEE, pp 454–459
47. Basheer IA, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 43(1):3–31
48. Kim T, Adali T (2002) Fully complex multi-layer perceptron network for nonlinear signal processing. J VLSI signal Process Syst Signal Image Video Technol 32(1–2):29–43
49. Hall JK (2019) The contributions of conversation analysis and interactional linguistics to a usage-based understanding of language: expanding the transdisciplinary framework. Mod Lang J 103:80–94
50. Petitjean C, González-Martínez E (2015) Laughing and smiling to manage trouble in French-language classroom interaction. Classroom Discourse 6(2):89–106
51. Pavlidis I, Tsiamyrtzis P, Shastri D, Wesley A, Zhou Y, Lindner P, Bass B (2012) Fast by nature-how stress patterns define human experience and performance in dexterous tasks. Sci Rep 2:305
52. Kosonogov V, De Zorzi L, Honoré J, Martínez-Velázquez ES, Nandrino JL, Martinez-Selva JM, Sequeira H (2017) Facial thermal variations: a new marker of emotional arousal. PLoS ONE 12(9):e0183592
53. Salazar-López E, Domínguez E, Ramos VJ, De la Fuente J, Meins A, Iborra O, Gómez-Milán E (2015) The mental and subjective

skin: emotion, empathy, feelings and thermography. Conscious Cogn 34:149–162
54. Wang S, Shen P, Liu Z (2012, October) Facial expression recognition from infrared thermal images using temperature difference by voting. In: 2012 IEEE 2nd international conference on cloud computing and intelligence systems, vol 1. IEEE, pp 94–98



**Chiara Filippini** obtained a master's degree in Biomedical Engineering at the University of Bologna in 2014. In 2017, she began her PhD program in the frame of Innovative PhD with industrial characterization program, at the Department of Neuroscience, University "G. D'Annunzio" Chieti-Pescara. During her PhD she was a visiting scholar at "Brain and Language Laboratory" of Gallaudet University in Washington DC, USA. Her main research area concern the study and modeling of the interaction between human and artificial agents. Aimed to increase the ability of the artificial agent to recognize and process the psychophysiological or cognitive state of the human interlocutor using is non-invasive neuroimaging techniques. Techniques and methodologies used are thermal infrared imaging and near-infrared functional spectroscopy.

**Edoardo Spadolini** obtained a master's degree in applied mathematics at La Sapienza University of Rome in 2017. His interests include software development, computer vision, human-machine interaction and embedded systems. He works in Research and Development at Next2U S.r.l.

**Daniela Cardone** obtained the master's degree in Biomedical Engineering at the La Sapienza University of Rome in 2009. In 2013, she obtained the PhD title in Neuroscience and Neuroimaging at the University "G. d'Annunzio "of Chieti-Pescara. From March 2020 he is RTD-A for the PON ADAS + project "Development of advanced technologies and systems for car safety through ADAS platforms" at the Department of Neuroscience, Imaging and Clinical Sciences at the University of "G. d'Annunzio "of Chieti-Pescara. Her research work mainly concerns the development of processing methods and analysis of physiological images and signals. More recently, her research focused on affective computing and human-machine interaction, with particular reference to automotive.

**Domenico Bianchi** is a R&D Engineer and Innovation Manager at Ud'Anet. His interests are in artificial intelligent systems, control theory and its applications, in automotive, industrial, traffic and robotics systems. He received Bachelor of Engineering degree and Master of Engineering degree in Control Systems and Computer Science, from the University of L'Aquila, in 2005 and 2008, respectively. In March 2012' he got Doctoral Degree (PhD) in Electrical and Information Engineering from the University of L'Aquila. From 2012 to 2015, he was appointed as a Post-Doctoral Researcher at the Centre of Excellence for Research DEWS, University of L'Aquila on control of networked systems.

**Maurizio Preziuso** is Graduated in Engineering, for over twenty years he has been dealing with feasibility studies, designing IT systems and industrial and process automation. Currently he is in charge of the technical direction of the technology company Ud'Anet, owned by the University of Chieti-Pescara. He plays the role of project manager in numerous European research projects and has skills in the development of software and applications of augmented reality and artificial intelligence.

**Christian Sciarretta** is an entrepreneur working for over twenty years in the ICT sector. After a consultancy period in 2005 he founded Ud'Anet s.r.l. The company offers consulting and research services in the ICT sector and promotes technology transfer actions for innovation in learning.

**Valentina del Cimmuto** obtained a master Degree in educational sciences. She is currently working as a project manager at Ud'Anet s.r.l, managing and evaluating national and european projects and training courses, organizing seminars and congresses for the dissemination of good practices, elaborating proposals for local, national and european projects.

**Davide Lisciani** is the Chief Executive Officer and Chief R&D Officer at Liscianigiochi. With a degree in Electrical Engineering, he studied in the Quantistic Electronic Institute of CNR working on laser probe systems for photoablation lasers. Has got many years of experience in managing of R&D teams for industrial purposes.

**Arcangelo Merla** Ph.D., is Professor of Applied Physics at the Department of Neuroscience, Imaging and Clinical Sciences, at the G. d'Annunzio University of Chieti-Pescara (Italy), and Director of the Infrared Imaging Laboratory at the ITAB - Advanced Biomedical Technologies Institute, G. d'Annunzio University of Chieti-Pescara. In 2018, he received the Distinguished Visiting Professor Award at the "Brain and Language Laboratory for Neuroimaging" - Gallaudet University, Washington DC, USA. He is founder of Next2U s.r.l.(www.next2u-solutions.com), an ITC company which develops systems for computational psychophysiology for human-machine interaction. Prof. Merla's recent research activity is focused on affective computing and human- machine interaction, with particular reference to assistive robotics, smart cars and assisted ambient living.