A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI

Derek McColl · Alexander Hong · Naoaki Hatakeyama · Goldie Nejat · Beno Benhabib

Received: 10 December 2014 / Accepted: 11 August 2015 / Published online: 23 August 2015

Abstract In Human-Robot Interactions (HRI), robots should be socially intelligent. They should be able to respond appropriately to human affective and social cues in order to effectively engage in bi-directional communications. Social intelligence would allow a robot to relate to, understand, and interact and share information with people in real-world human-centered environments. This survey paper presents an encompassing review of existing automated affect recognition and classification systems for social robots engaged in various HRI settings. Human-affect detection from facial expressions, body language, voice, and physiological signals are investigated, as well as from a combination of the aforementioned modes. The automated systems are described by their corresponding robotic and HRI applications, the sensors they employ, and the feature detection techniques and affect classification strategies utilized. This paper also discusses pertinent future research directions for promoting the development of socially intelligent robots capable of recognizing, classifying and responding to human affective states during real-time HRI.

Keywords Human-robot interactions · Affect classification models · Automated affect detection · Facial expressions · Body language · Voice · Physiological signals · Multi-modal

1 Introduction

Human-robot interaction (HRI) encompasses both the development of robots that engage humans in specific activities and tasks, as well as the study of how humans and robots interact with each other during such scenarios [1]. HRI can take many forms, from an operator teleoperating mobile robots during search and rescue operations [2], to robotic office-building mail delivery [3], to a socially engaging nurse robot delivering medicine to elderly long-term care residents [4].

A. Hong · B. Benhabib (✉)
Computer Integrated Manufacturing Laboratory,
Department of Mechanical and Industrial Engineering,
University of Toronto, 5 King's College Road,
M5S 3G8 Toronto, Canada
e-mail: benhabib@mie.utoronto.ca

A. Hong
e-mail: alex.hong@mail.utoronto.ca

D. McColl · A. Hong · N. Hatakeyama · G. Nejat
Autonomous Systems and Biomechatronics Laboratory,
Department of Mechanical and Industrial Engineering,
University of Toronto, Toronto, Canada

D. McColl
e-mail: derek.mccoll@mail.utoronto.ca

N. Hatakeyama
e-mail: naoaki.hatakeyama@mail.utoronto.ca

G. Nejat
e-mail: nejat@mie.utoronto.ca

Past HRI research has mainly concentrated on developing efficient ways for humans to control/supervise robot behavior [1]. More recently, research has also focused on developing robots that can detect common human communication cues for more natural interactions [5–7]. Social HRI is a subset of HRI that encompasses robots which interact using natural human communication modalities, including speech, body language and facial expressions [8]. This allows humans to interact with robots without any extensive prior training, permitting desired tasks to be completed more quickly and requiring less work to be performed by the user [8]. However, in order to accurately interpret and appropriately respond to these natural human communication cues, robots must be able to determine the social context of such cues.

For robots to successfully take part in bi-directional social interactions with people, they must be capable of recognizing, interpreting and responding effectively to social cues displayed by a human interactant. In particular, during social interactions, a person's affective displays can communicate his/her thoughts and social intent [9]. A person's affect is a complex combination of emotions, moods, interpersonal stances, attitudes, and personality traits that influence his/her own behavior [10]. A robot which is capable of interpreting affect will have an enhanced capacity to make decisions and assist humans due to its ability to respond to their affect [11]. Such robots would promote more effective and engaging interactions with users that would lead to better acceptance of these robots by their intended users [12].

Affective interactions between a human and a robot may take several forms, including: collaborative HRI [13], assistive HRI [14], mimicry HRI [15], and general HRI (e.g., multi-purpose) [16]. Collaborative HRI involves a robot and a person working together to complete a common task [17, 18]. A robot needs to be able to identify a person's affective state to avoid misinterpreting social cues during collaboration and to improve team performance [19]. Assistive HRI involves robots that are designed to aid people through physical, social and/or cognitive assistance [20, 21]. Automated affect detection would promote effective engagement in interactions aimed at improving a person's health and wellbeing, e.g. interventions for children with autism [22] and for the elderly [23]. Mimicry HRI consists of a robot or person imitating the verbal and/or nonverbal behaviors of the other [24]. Mimicry of affect can also be used to build social coordination between interacting partners and encourage cooperation [25]. Lastly, general or multi-purpose HRI involve robots designed to engage people using bi-directional communication for various applications.

Previous surveys on automated affect recognition have investigated approaches that utilized to determine affect from facial expressions [26–28], body language [28, 29], voice [27, 28] or a combination of facial expressions and voice [27]. In [28], additional modalities including physiological signals and text were also investigated, as well as various combinations of two or more of the aforementioned modalities [28]. These surveys primarily focused on Human-Computer Interactions (HCI) applications, as opposed to HRI addressed in this paper. HCI is the design, planning and study of devices and systems that allow or improve how a human interacts with virtual objects or agents on a computer [30]. HRI, on the other hand, requires an embodied physical robot to interact with a person or multiple persons. Embodied social robots have been designed to display human-like social behaviors in order to encourage humans to engage in social interaction with the robot, e.g. [31, 32]. A number of studies have shown that physically-embodied agents offer more engaging, enjoyable and effective social interactions than do computers and/or virtual agents [33–35].

A recent survey on perception systems for social HRI was presented in [36]. This survey focused on discussing perception methods used by robots mainly for face and object recognition, localization and tracking. Systems developed to allow social robots to autonomously identify a person's affect require specific sensors that will not interfere with or influence a person's behaviors during the desired HRI scenario. Socially interacting robots should also be able to interpret and respond to natural (non-acted) displays of affect. Within this context, it is important to investigate the desired HRI scenarios and as well the affect feature detection and classification systems used by robots to recognize and interpret human affect. Furthermore, the real-world experimental scenarios that have been studied and their corresponding results should be examined to provide insight into this emerging research area.

Uniquely, the objective of this paper is to present the first extensive survey of automated affect recognition and classification techniques used by social robots in varying HRI scenarios. The contributions of the paper are summarized as follows:

1. Descriptions of most common modalities utilized to detect affect and corresponding categorization models, namely: facial expressions, body language, voice, physiological signals, and multimodal techniques.
2. Classification of each affect-detection system with respect to its proposed HRI scenario, including collaboration tasks between humans and robots, assistive tasks, where robots aid a human to complete, mimicry tasks, where robots/humans perform the same behavior, and multi-purpose tasks that can potentially encompass a large range of HRI activities.
3. Descriptions of affect-detection systems, focusing on their integration and use by social robots, their perception and feature identification techniques, their affect models and classification methods, and direct results from HRI experiments.
4. Discussion of current challenges and comparisons of existing approaches as well as future research directions.

The remaining sections of this paper are organized as follows: Section 2 reviews the affect categorization models utilized for automated affect detection. Sections 3, 4, 5, 6 and 7 present affect recognition and classification systems for social robots focusing on affect detection from facial expressions, body language, voice, physiological signals, and a combination of the aforementioned modes, respectively. Section 8 presents a discussion of the research problems and approaches that have been addressed in the literature, and potential future trends. Section 9 provides concluding remarks.

2 Affect Categorization Models

The neuroscience and cognitive science fields have identified two leading types of models that describe how people perceive and classify affect: categorical and dimensional [37]. In general, categorical models consist of a finite number of discrete states, each representing a specific affect category [37, 38]. For example, a form of affective expression such as a facial expression, body language display, vocal intonation or a combination of these can be classified into one of these states: happy, sad, angry, etc. Dimensional models, on the other hand, also known as *continuous* models, use feature vectors to represent affective expressions in multi-dimensional spaces, allowing for the representation of varying intensities of affect [37, 39]. Changes in affect are considered in the continuous spectrum for these models. This section discusses existing variations of these two leading models and how they have been used by robots in HRI scenarios.

2.1 Categorical Models

In 1872, Darwin first classified six universal human affective states that were cross-culturally innate by studying expressions and movements in both humans and animals [40]. These six states were defined to be happiness, surprise, fear, disgust, anger, and sadness. In 1962, Tomkins also proposed that there were a limited number of human affective states by observing human communication movements such as nods and smiles [7]. He introduced the Affect Theory, which included nine discrete affect categories: joy, interest, surprise, anger, fear, shame, disgust, dis-smell, and anguish [41, 42]. Similar to Darwin, Ekman also conducted human cross-cultural studies and determined that human facial affect was a collection of the same six recognizable basic categories [43]. Ekman codified the corresponding facial expressions into these affective states using the Facial Action Coding System (FACS), where positions of facial action units (AU) are classified as distinct facial expressions of emotions [44]. This categorical model is debatably the most often used model today for classifying affect [45].

The strength in categorical models lies in their ability to clearly distinguish one affect from the others. However, categorical models lack the ability to classify any affect that is not included as a category in the model.

2.2 Dimensional Models

In order to address the challenge of classifying affect across a continuous spectrum, dimensional models have been developed. Russell [39] and Barrett [46] argued that affect cannot be separately classified as gradations, and blends of affective responses could not be categorized. Furthermore, they argued that there were no correspondences between discrete affect and brain activity. The majority of dimensional models that have been developed use either two or three dimensions for affect categorization.

In 1897, Wundt first proposed that affect could be described by three dimensions: pleasure *vs*. displeasure, arousal *vs*. calm, and relaxation *vs*. tension [47]. The pleasure-displeasure dimension described the positivity or negativity of mental state, the arousal-calm dimension referred to excitation level, and the relaxation-tension dimension referred to the frequency of affect change [47]. In 1954, Schlosberg proposed a similar model also consisting of three dimensions of affect: pleasure *vs*. displeasure, attention *vs*. rejection, and level of activation (also known as relaxation *vs*. tension) [48]. The attention-rejection dimension described the levels of openness of a person [49].

In 1980, Plutchik designed a three-dimensional model which included valence, arousal, and power. The model combined both categorical and dimensional model theories as it incorporated blends of eight basic affective states: curiosity, fear, disgust, sadness, acceptance, joy, anger, and surprise [50]. The valence dimension described how positive or negative an affect was, similar to that of the abovementioned pleasure-displeasure scale, and the power dimension described sense of control over the affect.

In [51], Mehrabian suggested the pleasure, arousal, and dominance (PAD) emotional state to incorporate all emotions. Dominance described the dominant or submissive nature of the affect. Russell also proposed a two-dimensional circumplex model containing valence and arousal dimensions [52]. The model illustrated that affect was arranged in a circular fashion around the origin of the two-dimensional scale. The circumplex model has been widely used to evaluate facial expressions [53].

Watson proposed a two-dimensional PANA (Positive Activation-Negative Activation) model to better account for affect consisting of high arousal and neutral valence [54]. The PANA model suggests that positive affect and negative affect are two separate systems, where states of very high arousal are defined by their valence and states of very low arousal tend to be more neutral in valence [55].

Dimensional models have the ability to encompass all possible affective states and their variations [56]. However, they lack the strength to distinguish one prominent affect from another and can often be confused with each other.

2.3 Affect Models Used in HRI

A number of HRI studies have incorporated the use of affect classification using categorical emotional models for facial expressions [57–60], body language [61–64], voice [7, 19, 65–69], physiological signals [70–72], and multi-modal systems [73, 74]. These models allow robots to interpret affective states in a similar manner as humans [75]. The most common discrete affective categories used by robots in HRI settings are disgust, sad, surprise, anger, disgust, fear, sad, happy, surprise, as well as neutral. These affective categories encompass Ekman's six basic affect and have been used to infer the appropriate social robot response in various HRI scenarios.

HRI studies have also been conducted with robots using dimensional models for affect classification using facial expressions [76–78], body language [5, 79, 80], voice [81, 82], physiological signals [4, 22, 70, 83–87], and multi-modal systems [88–90]. The most common model used in HRI is the two-dimensional circumplex (valence-arousal) model. This model captures a wide range of positive and negative affect encountered in common HRI scenarios.

HRI researchers have also developed their own dimensional affective models. For example, in [91], a four-dimensional model was developed for multimodal HRI to determine affect from voice, body movements, and music. The dimensions of the model were speed, intensity, regularity, and extent. The use of both categorical or dimensional models has allowed social robots to effectively determine a person's affect and in turn respond to it by changing its own behavior which has also included the display of the robot's own affect.

3 Facial Affect Recognition During HRI

Facial expressions involve the motions and positions of a combination of facial features, which together provide an immediate display of affect [92]. Darwin was one of the first scientists to recognize that facial expressions are an immediate means for humans to communicate their opinions, intentions, and emotions [40]. Schiano et al. also identified facial expressions to be the primary mode of communication of affective information due to the inherently natural face-to-face communication in human-human interactions [93]. Fridlund emphasized the cross-cultural universality of emotions displayed by the face [94], and that facial expressions are used for social motives [95], and may directly influence the behaviors of others in social settings [96].

A key feature to social robots is their ability to interpret and recognize facial affect in order to have empathetic interactions with humans [97]. With the advancement of computer-vision technologies, there has been significant research effort towards the development of facial-affect-recognition systems to promote robot social intelligence and empathy during HRI [97]. In order to emulate empathy, robots should be capable of recognizing human affective states, displaying their own affect, and expressing their perspective intentions to others [98]. There are several challenges in recognizing human facial affect during HRI including: lack of control over environmental lighting conditions and working distances [58, 98], real-time computational requirements [58, 99], processing spontaneous dynamic affective expressions and states [100], and physical and technology constraints [101].

The majority of facial affect recognition systems use 2D onboard cameras to detect facial features. Facial-affect classification is, then, used to estimate affect utilizing binary decision trees [58], 2) AdaBoost [102], multilayer perceptrons (MLPs) [60], support vector machines (SVMs) [103], support vector regression (SVRs) [60, 76], neural networks (NNs) [60, 104–108], or dynamic Bayesian networks (DBNs) [75].

This section presents a detailed review of the state-of-the-art human facial affect recognition systems for HRI scenarios.

3.1 Collaborative HRI

A collaborative HRI scenario was presented in [78, 88], where a cat-like robot, iCat, with an expressive face was used to play chess on an electronic chessboard with children and provided real-time affective feedback to their chess moves. The study aimed to investigate how children perceive and collaborate with an empathic social robot in a school environment. During game playing, a child's affect was captured by a 2D webcam and categorized as positive, neutral or negative valence. The inputs into the affect model were behavioural and contextual features, which included the probability of a smile, eye gaze, game state and game evolution. The probability of a smile was determined using SVMs and facial geometric features obtained from both 2D and 3D facial landmarks extracted using the Seeing Machine's faceAPI [109]. Eye gaze, which was defined by the pose of the head relative to the webcam, was also tracked using faceAPI. The game state and game evolution between players were determined by the configuration of the chess game [110]. The affect-detection system, then, used these features as inputs into SVMs in order to estimate the probability of each of the three valence levels. The overall system was trained with features from the Inter-ACT affective multi-modal video corpus [111], which contained videos of interactions of children playing chess with iCat. Based on the valence level, iCat employed an empathic strategy when the player was not experiencing positive valence with high probability by either providing encouraging comments, scaffolding by giving feedback on a user's previous move, offering help by suggesting a good move, or making a bad move on purpose. A reinforcement learning technique was used to adapt the robot's behavior over time to a player. Experiments were conducted with 40 elementary school children playing chess with the iCat in one of three control groups, where iCat had no empathic response, had random empathic response, or had adaptive empathic response. After the interaction, each child was asked a number of questions in an interview setting to evaluate the empathic performance of iCat. All children in the adaptive empathic control group believed that iCat knew how they were feeling compared to 77 % of the children from the random empathic control group, and 50 % of the children from the neutral control group. The study concluded that empathy was an important mechanism for robots to have in order to facilitate interactions with humans.

In [77], the humanoid Nao robot played a question-based quiz game on a tablet with a child while they were seated across from one another. Both players would take turns asking and answering questions that appeared on the tablet. The robot's emotions were influenced by both the emotions of the child and the state of the game. The emotional state of the child was determined by the human operator tele-operating the robot in a Wizard of Oz manner. Namely, the human operator determined the arousal and valence values for the children following guidelines based on the location of specific facial emotions (e.g., angry, excited, calm, bored, sad) on a 2D affect model. These values were, then, used as inputs along with the robot's dialogue into the Nao robot to model its own affective state. Facial emotions were recorded using a separate 2D environment video camera. Experiments consisted of investigating the interactions of two Nao robots, one which adapted its behavior to the emotions of the child and one that did not. The robots played the quiz game with 18 child participants in a counter-balanced procedure. The level of extroversion of the child on a scale of 0 to 100 was determined beforehand using the BFQ-C questionnaire [112]. This value was used to select the speech volume and gesture size of the robots during interactions. Questionnaires regarding the opinions of children with respect to both robots, and video analysis of the interactions found that the children had more expressions and reacted more positively towards the affective robot than the non-affective robot. The study showed that a robot with adapting expressive behaviors could influence the expressive behaviors of children and engage them in a collaborative HRI.

3.2 Assistive HRI

In [58], the B21r robot utilizing an onboard 2D camera was used to identify a user's facial affect for the intended application of assistive learning. The robot had an LCD display to show its own virtual face and facial expressions. As a teacher, B21r could use facial affect information to evaluate the quality of the class lessons it was providing and progress from one lesson to another. The facial expression identification approach consisted of extracting skin-color information from images provided by the 2D camera, then, localizing the entire human face and corresponding facial components and, lastly, fitting a deformable model to the face using a learning-based objective function. Structural and temporal facial features were used as inputs to a binary-decision tree in order to classify the following seven affective states: anger, sad, happy, disgust, fear, neutral, and surprise. Experiments with the facial-affect-classification system using the Cohn-Kanade facial expression database [113] obtained a recognition rate of 70 % based on 488 image sequences of 97 individuals displaying the aforementioned facial expressions. A preliminary experiment was also conducted with the B21r robot, where a subject stood facing the robot and displayed neutral, happy and surprise facial expressions. 120 readings of each expression resulted in a mean recognition rate of 67 %.

In [102], an autonomous mobile service robot, named Johnny, was designed to serve guests in a restaurant. The robot was capable of receiving seat reservations, welcoming and escorting guests to seats, taking and delivering an order, and entertaining guests by playing appropriate songs based on their affect. The robot was equipped with pan-tilt stereo cameras, which were used to obtain still images of the following seven facial expressions: sadness, surprise, joy, fear, disgust, anger, and neutral. Gabor filters were used to extract texture and shape information and detect local lines and edges for each facial expression. The multi-class AdaBoost.MH learning method [114] was used for Gabor feature selection The AdaBoost.MH algorithm was trained using randomly created training sets of images from the Cohn-Kanade database [113] to find which Gabor features were the most relevant to differentiate between different facial expressions. Two weak learners were analyzed for the AdaBoost.MH: multi-threshold and single-threshold decision stumps. The single threshold decision stump aimed to find a single threshold that best separated all facial expressions, and the multi-threshold version found multiple thresholds, one threshold for each facial expression, that best separated all facial expressions. The facial-recognition system was also tested on the Cohn-Kanade database, where recognition experiments were repeated five times on randomly generated training and test sets from the database. It was found that the multi-threshold weak learner performed better due to its average error rate being lower, where the minimum average error rate obtained using 900 features was 9.73 % for single-threshold decision stumps, and 8.84 % for multi-threshold decision stumps using only 500 features. In addition to having facial expression recognition capability, the robot was also able to manipulate objects with its hand to deliver restaurant orders, plan its path and localize in the environment, and recognize both speech and hand gestures from guests as demonstrated in a RoboCup@Home [115] competition.

There have also been a number of automated facial-affect-detection systems that have been developed for assistive HRI scenarios but that have not yet been implemented and tested on robotic platforms [57, 59]. One of the earlier works can be found in [57], where a human affect-recognition system was developed for the purpose of robots carrying out intelligent conversations and for the application of better living. 60 feature points from the eyebrows, eyes, and mouth were obtained from images provided by a CCD camera. Twenty-one additional parameters were also generated from the 60 feature points. These parameters expressed the shapes of the eye, eyebrow and mouth. The feature parameters were used as inputs into a back propagation NN to recognize the six affective states of surprise, fear, disgust, anger, happiness, and sadness. Experiments with 30 clients resulted in recognition rates ranging from 66.4 % to 91.2 %. The results showed the potential of the system to be used in robotic applications.

3.3 Mimicry HRI

Another form of HRI is mimicry, where the robot imitates the facial expressions of a human user. In [116], the Sony AIBO robotic dog was used to imitate the facial affect of a person by using an array of LEDs on its own face, and by moving its body, head and ears. In order for the robot to mimic human affect, head and face tracking [117] was used to compute the facial feature positions and 3D head pose of a real-time video sequence of human facial expressions obtained from an environmental camera. The facial-affect-recognition approach consisted of capturing the temporal pattern of facial features, namely, the lip and eyebrows during the video sequence to detect when facial actions changed abruptly. A temporal classifier was, then, used to classify facial expressions into the five following emotions: surprise, sad, joy, disgust, and anger. The sequence of facial actions was modeled by a second-order auto-regressive function and frames of interest were detected by identifying a local positive maximum in the temporal derivatives of a facial action. A performance study on this method was done on 35 test videos featuring seven people from the Carnegie Mellon University (CMU) database [118] and recognition rates were 80 % in classifying facial expressions. Subsequently, the AIBO robot was used to imitate the expressions of a human subject using a 1600-frame sequence of human expressions. The results showed that AIBO was able to mimic human expression with only a slight delay.

In [60], the iCat robot was used in order to have people imitate its facial expressions. A 2D camera in the robot's nose was used to obtain images of the person. A semi-automated technique was used to find the image region that contained a face. To reduce the number of features to extract from an image, features were extracted using one of three techniques: Independent Component Analysis (ICA) [119], Principal Component Analysis (PCA) [119], and Non-negative matrix Factorization (NMF) [120]. Three regression techniques were, then, compared for direct facial expression imitation: SVR [121], MLPs [122], and cascade-correlation NNs [122]. Training data from 32 subjects imitating 50 facial expressions (i.e., fear, anger, surprise, happiness, disgust, and sadness) of the robot under two different lighting conditions were used. Experiments consisted of capturing a maximum of 100 facial images for each of the 32 subjects. A five-fold cross-validation approach was utilized to determine how often each regression estimator resulted in successfully imitating a facial expression. The SVR approach had the most accurate results with and without the feature-extraction stage.

In both [104] and [105], a robotic head was developed to recognize human facial affect through an imitation game with humans in order to investigate emotional interaction-based learning. The robotic head consisted of two eyes (each having a 2D camera), two eyebrows, and a mouth. The robot's state of mind was considered to be a "baby", where the robot knew very little about its surroundings and learned through interaction, and a human was considered to be a parental figure. The facial-affect states identified during the interactions were happiness, sadness, anger, surprise, and neutral. Online learning of the facial expressions was achieved using a NN-based architecture. Facial expressions were learned without the need to locate and frame a face. Focus points, such as eyes, eyebrows, nose, and mouth, were used to recognize facial expressions and infer the location of the face. These focus points were learned through an unsupervised pattern matching strategy called Self Adaptive Winner takes all (SAW) [123], and a combination of log-polar transform and Gabor filters were applied to extract visual features. In [104], during the learning phase, ten different humans were asked to imitate the facial expressions of the robotic head in random order. The robot learned these facial expressions though a sequential exploration of the focus points in 1,600 images taken by one of its 2D cameras and attempted to condition them to a given facial expression. Experiments consisting of 20 new people interacting with the robot were conducted to test the generalization of the approach [104]. Results of a 20-fold cross validation indicated that the robot could imitate persons not included in the learning phase. In [105], an additional experiment was presented to investigate the success rate for the facial identification system using the same 20 people database mentioned above. Three feature-extraction techniques were compared including using log-polar transform, Gabor filters, or a combination of both. The results showed that the average success rate was better for all emotions when using the hybrid feature-extraction approach, except for sadness. Sadness resulted in low recognition as each person imitating the robot showed sadness in a different manner believed to be as a result of the context of the situation.

An animated face robot was developed in [106] to produce human-like facial expressions and also recognize human facial affect. Human facial affect was recognized using facial images obtained by a CCD camera mounted inside the robot's left eye. The monochrome CCD camera captured brightness distribution data of a face and this information was used to locate the positions of each iris of a person's eyes. Then 13 vertical lines were selected that crossed facial characteristic points in each area of the face, i.e., the eyebrows, eyes, and mouth. These two types of information were used together as inputs into a NN for facial expression recognition. The NN was trained through back propagation using a video of 15 subjects changing their facial expressions from neutral to each of the following six expressions: anger, surprise, disgust, fear, sadness and happiness. The offline trained NN was, then, used in recognition tests with a video of another 15 subjects. The recognition rate was determined to be 85 % at a rate of 55 ms.

In [124], an experiment was performed with the animated face robot to examine the personality of the robot (i.e., response behavior to a person). In the experiment, a Q-learning algorithm was implemented to enable reinforcement learning of the robot's preferable responses. Three types of rewarding were utilized: AND, OR and COMPARATIVE. Five subjects provided positive, negative or neutral rewards based on the ability of the robot face to express happiness. The results showed that both the OR and COMPARATIVE rewarding types were effective in the learning process and can be used for a robot to learn its own appropriate facial expression during communicative interaction with a person.

In [125], a KANSEI communication system was developed for robot emotional synchronization with a person. KANSEI encompasses emotion, intelligence, feeling, impression, sensitivity, and intuition to promote natural communication [126], The head-robot KAMIN was used to recognize human affect through facial expressions and, then, synchronize its own affective state. KAMIN had its own facial expressions projected onto its dome screen using a fish-eye lens projector. Images from a 2D video camera were used for face detection and tracking using AdaBoost [127]. Namely, AdaBoost was used to detect critical facial features, such as eyes, mouth, and eyebrows. Facial expression recognition was performed by using a 2D continuous scale with bending and inclination as its two dimensions [128]. The scale was based on hand-coded expressions of 43 subjects displaying happy, angry, fear, sad, disgust and surprise. Experiments were conducted with 12 university students displaying the aforementioned six emotions in random order. In order to induce these emotions, the students were shown pictures and videos and provided with selected topics. The bending and inclination parameters were, then, obtained and plotted in the 2D emotional recognition space. The average recognition rate was determined to be 77 %. In order for the robot to display the synchronized emotions, a vector field of dynamics was designed into the 2D circumplex (valence-arousal) model. Namely, the identified human emotion was mapped onto the model and the vector field was used to direct the robot's emotion towards where the human emotion was mapped in the 2D space. Forty university students were used to test the synchronization; twenty interacted with KAMIN when it had synchronized emotions, and twenty interacted with the robot when it did not have synchronized emotions. The robot was able to match the same arousal and valence levels as the subjects during the interactions and the subjects were more willing to communicate with the robot when it synchronized with their emotions.

In [75], a real-time system for recognition and imitation of facial expressions was developed for the robotic head Muecas. Muecas was equipped with an onboard camera to capture images of a person's face for facial expression recognition. The five facial expressions recognized were: fear, happiness, anger, sadness, and neutral. Facial expression recognition was accomplished in three steps: face detection, invariant facial feature extraction, and facial expression classification. Face detection was achieved using the Viola and Jones face detection method [129]. Noise and illumination effects were improved by using a series of filters, including a Gabor filter for easier extraction of edge-based facial features. Invariant edge-based features were then extracted using a set of AUs. The features points were selected based on FACS [130]. These features were, then, used as inputs into a DBN classifier, which was used to categorize facial expressions into the 5 emotions. Offline training was accomplished by fitting a Gaussian distribution to learning data. Imitation was achieved through the mapping of the AUs to a mesh model of Muecas prior to direct mapping to the robot's actuators to avoid any possible collisions when generating expressions. Experiments were conducted with 20 participants performing the facial expressions randomly five times. Results showed a detection rate of over 90 % for all facial expressions.

3.4 Multi-Purpose HRI

In [76], a facial-affect-detection system was developed for social HRI settings. The proposed methodology utilized 2D images extracted from a video camera onboard a robot to identify facial feature points. These feature points were then converted into a unique set of geometric-based Facial Expression Parameters (FEPs). The non-dimensionalized FEPs allowed for affect estimation from any facial expression without the need for prior expressions as a baseline. SVR was, then, used to determine the valence and arousal levels corresponding to spontaneous facial expressions using the FEPs as inputs. Experiments with the JAFFE database [131], and a database consisting of images of individuals displaying natural emotions created by the authors of [76] showed affect-prediction rates from 74 to 94 % and 70 to 72 % for valence and arousal, respectively, at a 15 % tolerance level. Preliminary HRI experiments were also conducted consisting of one-on-one interaction scenarios between the expressive social robot Brian 2.0 and five participants during two different types of interaction stages: a getting-to-know you stage and a storytelling stage. At a 20 % tolerance level, valence and arousal prediction rates were determined to be 72 % and 70 %.

In [132], the humanoid robot, RObot huMan interAction machiNe (ROMAN), utilizing a mounted 2D camera on its body was used to recognize human emotions. The affect of the robot itself was influenced by the affect of the person during interaction. For example, the robot would respond slower and show empathy when the person was in a state of sadness. A Haar cascade classifier and a contrast filter were used to localize a person's facial features, including the pupils, eyebrow, mouth, and lower and upper lip. These feature points were selected based on FACS [130]. Based on the exact feature point locations and measurement of action unit activities from FACS, a probability was associated with the six basic emotions: disgust, sadness, fear, happiness, surprise, and anger. Experiments consisted of testing emotion interpretation and emotion detection on a data set of the six basic emotions displayed by different faces [132]. Results showed a recognition rate of about 60 % for the emotions of surprise, sadness, and happiness.

In [103], the social mobile robot RobotVie was used to recognize human-facial expressions in unconstrained environments. The robot used four environmental video cameras positioned at different angles for processing human facial expressions during face-to-face interaction in real-time and was capable of categorizing facial expressions into seven affective states: surprise, joy, neutral, sadness, anger, disgust, and fear. Face detection was achieved via video data using an extended version of the Viola and Jones face-detection method [129]. Facial expression classification was implemented using a combination of SVM and AdaBoost learning techniques, AdaSVM. Namely, AdaBoost was used to extract a reduced set of Gabor features to use for training the SVM. Initial training and testing of the affect-classification method was conducted using the Cohn-Kanade database [113]. The technique was also evaluated with the Pictures of Facial Affect database [133]. Recognition results for the two databases were 93 % and 97 %, respectively. Moreover, a HRI experiment was conducted with 14 participants displaying facial expressions to RoboVie. The participants were instructed to have both spontaneous head movements and facial expressions. To test the affect-recognition system, the recorded video was slowed down to 1/3 speed and four human observers adjusted a dial to indicate the amount of happiness they observed in each video frame. A 28-dimensional vector was defined for each video frame based on the output of the classifiers for each of the 7 emotions and the 4 cameras. An average correlation of 0.56 was found between the observers and the robot on training data, and 0.30 on new images of the same 14 participants.

In [134], during an HRI scenario, the expressive Einstein robotic head was used to directly respond to the affect of multiple people during HRI. The robot was able to recognize the facial expressions of multiple people simultaneously, and speak and display its own facial expressions to the group. An environmental 2D camera situated in front of the robot was used to capture the faces of the multiple people. The robot classified their facial expressions into seven classes including fear, happy, anger, sad, disgust, neutral, and surprise. Facial expression recognition was accomplished by combining and statistically weighting two classification methods. Firstly, facial features were extracted using Active Appearance Models (AMM), where shape and texture information were used to identify feature points on the face, such as eyes, mouth and eyebrows. Facial expressions were then classified using a features vectors- based approach, using the distances and ratios of these points. Secondly, differential AAM features were determined using the difference of AAM features between a displayed facial expression and a neutral face using a low dimensional manifold space. The two approaches were using weights in a linear discriminate function. This method was applied to each face detected in the scene and the weight of each face was determined by its size, namely, the closer to the robot, the more dominant was the person's emotion on the overall emotion atmosphere. Experiments were performed on the Cohn-Kanade database [113]. The results showed a recognition rate greater than 90 % for happy, angry, and surprise. However for the negative emotions, the approach did not perform well due to the minimal distance differences between feature points. HRI experiments were also conducted with a group of three students interacting with the robot. The robot was able to determine the affective atmosphere obtained from the facial expressions of the students and displayed its own facial expressions and speech.

In [135] and [136], a social interaction model was developed for the AIBO dog robot to permit a robot and user to engage in bi-directional affective communication. The AIBO detected the user's facial affect and responded with its own affective movements. Facial-affective states that were recognized by the robot included joy, sadness, surprise, anger, fear, disgust, and neutral. The dog robot was able to respond using emissions of sounds, lights, and motions. For instance, when the user displayed an angry face, the dog would display sadness by playing a sad song and lowering its head. An onboard CMOS camera located on the robot's head was used to obtain facial images. Facial expression recognition was accomplished using facial feature extraction and mapped onto the facial action units defined by FACS [130]. These facial movements were then mapped to facial expressions. A Probabilistic Finite State Automation (PFSA) method was used to determine the state of the robot. Q-learning was also applied to train the dog to respond in a certain way to a user's emotional state. In [136], experiments were conducted to motivate the AIBO robot to behave in a certain way. Through Q-learning, the robot adapted its behavior to the human's affective state. A friendly personality was achieved when the human interacted in positive states, such as joyful and happy.

In [107] and [108], a small humanoid robot, HOAP-3, was able to recognize human facial expressions and respond accordingly with a postural sequence of affective movements. Both AAMs and a Neural Evolution Algorithm were used for face shape recognition and facial expression classification. AAMs were used to identify facial features from the eyes, chin, nose, mouth, and eyebrows based on matching of a statistical model of appearance and shape variation to the facial image. The Neural Evolution Algorithm consisted of using a Differential Evolution (DE) optimizer with a NN system. The input parameters to the NN were the location of the facial features obtained using AAM. In [108], experiments were conducted evaluating the recognition rates for the happy, sad, neutral facial expressions in both static and real-time images of several different subjects. Sixty-three facial feature points were used. Recognition results for affective states ranged from 72.5 % for sad to 96 % for happy. Once recognized, the robot would implement a pre-programmed gesture based on the detected facial expression. Namely, it was able to respond with faster postural movements when the human was happy and slower movements when the human was sad.

A number of automated facial affect detection systems have also been proposed [97, 137–143], for general HRI scenarios, but have not yet been integrated on robots. For example, in [141], a fast learning algorithm that improved training time for SVMs used in facial affect classification was developed to be used by service robots. The approach involved using a CMOS image sensor to acquire facial images, and identify and extract facial features for affect recognition processing. Affect recognition was achieved by using Support Vector Pursuit Learning (SVPL) of the facial feature positions for different facial expressions. The method allowed automatic retraining of a new SVM classifier for new facial data by modifying parameters of the SVM hyperplane. Experiments consisted of evaluating the performance of both an SVM and SVPL on a dataset of 100 images from ten different individuals. The SVM was trained with 20 images and tested on the other 80 images. The recognition rate was recorded to be 86 %. In another experiment, SVPL was applied to retrain a SVM that was only recorded to have a 52 % recognition rate. With the new retrained SVM, the recognition rate increased to a 96 % recognition rate.

4 Body Language Based Affect Recognition During HRI

Body language conveys important information about a person's affect during social interaction [144]. Human behavioral research has identified a number of body postures and movements that can communicate affect during social interactions [145–147]. For example, in [145], Mehrabian showed that body orientation, trunk relaxation, and arm positions can communicate a person's attitude towards another person during a conversation. In [146], Montepare et al. found that the jerkiness, stiffness and smoothness of body language displays can be utilized to distinguish between happy, sad, neutral and angry affective states displayed during social interaction. Furthermore in [147], Wallbott identified specific combinations of trunk, arm and head postures and movements that corresponded to 14 affective states including the social emotions of guilt, pride, and shame. Body movements and postures have also been linked to dimensional models of affect [148, 149]. For example, it has been identified that there exists a strong link between a person's level of arousal and the speed of movements during a body language display [148] and that head position is important in distinguishing both valence and arousal [149].

To-date, the majority of research on identifying body language during HRI has concentrated on recognizing hand and arm gestures as input commands for a robot to complete specific tasks including navigating an environment [150–158], identifying object locations [159–161], and controlling the position and movements of a robot arm [162, 163]. A fewer number of automated affect detection systems have been developed to specifically identify body language displays from individuals engaged in HRI [5, 15, 61–64, 79, 80]. These affect from body language systems, which are discussed in this Section, have been designed for various HRI scenarios, including collaborative HRI [61], assistive HRI [5, 63, 79], mimicry [15, 62, 80], and multi-purpose HRI scenarios [64].

4.1 Collaborative HRI

In [61], the expressive cat-like robot iCat was designed to play a one-on-one chess game with children. The engagement of a child in a game was determined through his/her seated trunk pose. A 2D color camera mounted in the environment was utilized to obtain a lateral view of the child. Each child wore a green shirt during the interaction, in order for his/her trunk to be segmented from the background information via color segmentation. The following features of the trunk were utilized to identify engagement: body angle, slouch factor, quantity of motion, contraction index and meta features consisting of the derivatives of the aforementioned features over time. These features were, then, used as input for classifying a child as engaged or not engaged using learning techniques. Experiments were conducted with five 8-year old children. Baseline engagement was identified by three trained coders watching videos of the interactions. The aforementioned body language features were used with 63 different classifiers from Weka [164] in order to identify engagement. Classification results showed that the ADTree and OneR learning techniques obtained the highest recognition rate of 82 % in identifying engagement.

4.2 Assistive HRI

In [79], the human-like social robot, Brian 2.0, engaged users in both a memory training exercise and providing instructions to build a picnic table. An automated static body pose interpretation and classification system was developed to determine what level of accessibility (openness and rapport) a user feels towards Brian 2.0 during these interactions. 2D and depth information from a Kinect^{TM} sensor mounted on the robot's trunk was used to make a 3D upper body ellipsoid model of the person's trunk and arm poses during the interaction. These poses were, then, used to determine a user's level of accessibility towards the robot based on the Position Accessibility Scale (PAS) of the Nonverbal Interaction States Scale [165]. The PAS identifies accessibility on a 4-level scale based on trunk orientations with a finer scaling of 12 levels when the arm orientations are also considered. A person was determined to be more accessible towards the robot during social interactions when his/her upper and lower trunks and arms were oriented more towards the robot. 18 university students, aged 19–35, engaged in one-on-one social HRI experiments with Brian 2.0. The results showed an accessibility recognition rate of 89 % on the 4-level scale and 86 % on the 12-level scale when compared to the accessibility levels identified by an expert coder.

In [5], the newer version of the robot, Brian 2.1, was utilized to provide social assistance to elderly residents of a long-term care facility during meal eating. A Kinect^{TM} sensor, mounted on the robot's chest, was used to identify natural dynamic body language displays by the users during the meal in order to determine their affective valence and arousal levels. A 3D upper body model was used to track dynamic body displays of a user utilizing the depth data obtained from the Kinect^{TM} sensor. The depth data of the user was segmented using a Gaussian mixture model of the background, connected component analysis and head and shoulders contours. The upper body model was fit to the depth data of the user utilizing a dynamic Bayesian network with a ray-constrained iterative closest point measurement model [166]. The following body features were, then, extracted from the depth data: the bowing/stretching of the trunk, distance of the hands from the trunk, head position, forwards/backwards and vertical motion of the body, expansiveness of the body and speed of the body. These features were utilized as inputs into seven different learning techniques in order to classify the displayed valence and arousal levels. Experiments were conducted at a long-term care facility with eight elderly participants, aged 83–92, interacting one-on-one with Brian 2.1 during two lunch meals. The results found that the Radial Basis Function Network (RBFN) learning technique obtained the highest recognition rate for valence at 77.9 % and both the Adaboost with Naïve Bayes (ANB) base classifier and Random Forest techniques obtained the highest recognition rates for arousal at 93.6 %.

In [63], a human-like robot, Nadine, utilized a gesture understand HRI (GUHRI) system comprising an environmentally located Kinect^{TM} sensor and a CyberGlove worn by the user, to identify specific affective, communicative, and action gestures of the user during a teacher (robot) – student (user) interaction scenario. The skeleton joint locations estimated with the Microsoft Kinect^{TM} SDK [167] and the Kinect^{TM} sensor as well as the CyberGlove were utilized as features for gesture classification. An energy based large margin nearest-neighbor technique was utilized to classify the features into one of the affective gestures of confidence and praise as well as the following gestures: asking a question, objection, stop, success, shake hand, weakly agree, call, drink, read, or write. Experiments were conducted with 25 participants of different genders, body sizes and races. The participants were asked to generate each of the aforementioned gestures for the robot during a teacher-student interaction, with the robot responding to each gesture with a specific behavior, e.g. when the participant displayed an affective praising gesture, Nadine responded by saying “Thank you for your praise.” Random subsets of varying sizes of the participants' gestures were utilized for training the large margin nearest neighbor classifier. Utilizing a subset of 4 training samples for each gesture resulted in a recognition rate of 90 % while a subset of 14 training samples resulted in a recognition rate of 97 %.

4.3 Mimicry HRI

In [62], the humanoid robot Nao was designed to engage children in movement imitation games, where a child was asked to mimic the movements of the robot. The robot, then, recognized and gave feedback on the child's movements. The hand wave was used as the movement for the robot to recognize. A 2D camera in the robot's eye was utilized to track the hand using skin color and motion information in the captured video stream. The robot, then determined the energy average amplitude of acceleration and frequency of local acceleration maxima of a wave. Experiments consisted of an adult actor performing happy, angry, sad, and polite waves. It was observed that distinct affective states could be identified based on their acceleration-frequency matrix for the hand wave gesture. This procedure of identifying distinctive affective features from hand waves was also presented in [168] for multi-purpose HRI scenarios.

Several researchers also investigated how social robots influence the affective body language of a user during mimicry HRI [15, 80]. In [15], the Robovie robot was designed as a *partner robot* that could offer mental and communicational support to humans. The study investigated how robot and user body language displays compared to the subjective experience of the interaction. A post-interaction questionnaire was utilized to obtain participant experience using five adjective pairs: good-bad, kind-cruel, pretty-ugly, exciting-dull, likeable-unlikeable. Robot- and human body language displays were recorded utilizing an optical tracking system. Experiments with 26 university students talking and playing rock-paper-scissors with the robot found that synchronized eye contact and body language were correlated with positive questionnaire ratings, i.e., a participant's more positive ratings corresponded to more interaction time with synchronized eye contact and body language.

In [80], the Nao robot was utilized to investigate how the mood of the robot influenced the mood of a user during an imitation game where the user was tasked with mimicking the robot's movements, where these movements could range from easy to difficult. Experiments were conducted with 36 university students, aged 19 to 41, playing the imitation game with the robot. Participants rated both the valence and arousal levels of the robot and their own valence and arousal levels using self-assessment manikins after each game. The results showed that valence and arousal levels of the participants matched the valence and arousal levels of the robot during the easy game. For the more difficult game, the participants performed the task better when the robot displayed a negative valence compared to a positive valence while implementing its movements.

4.4 Multi-Purpose HRI

In [64], Kirin, a full sized humanoid robot, utilized two sensor modalities consisting of an environmentally located Kinect^{TM} to track a person's movements and 40 touch sensors on its body in order to autonomously identify a person's affective touching gestures, e.g. a person hugging the robot. Skeleton tracking of the user using the Microsoft Kinect^{TM} for Windows SDK [167] was utilized to determine the following body language features: the mean, standard deviation, maximum, minimum, first and third quartile and interquartile distances of the 3D positions of the head, torso, right upper and lower arms, and hands. These features and readings from the touch sensors were utilized as inputs for a either a SVM or a k-nearest neighbor (kNN) learning technique for classifying a gesture as one of 20 investigated affective touch gestures, such as hugging or kissing that show affection and slapping or pushing that show aggression. Experiments were conducted with 17 participants (eight Japanese and nine non-Japanese, with an average age of 32). Each participant performed each of the 20 affective touch gestures. Results found that SVM obtained the highest affective gesture recognition rate of 90.5 %, while kNN had a recognition rate of 82.4 %.

A number of automated affective body language perception and identification systems have also been proposed in the literature for multi-purpose robotic use, but have not yet been integrated onto robots for social HRI [169–172]. For example, in [172] the Microsoft Kinect^{TM} sensor and its SDK [167] were utilized to identify body language features using Laban Movement Analysis in order to determine if a body language display could be recognized as a rejoicing or lamenting affective state.

5 Voice Based Affect Recognition During HRI

During social interactions, people can communicate affect with their voices [173]. Changes in a person's affect can result in changes to the position of his/her larynx, vocal fold tension and position and breathing rate as well as positions and shapes of the lips and mouth muscles, which all influence the tone and quality of the voice [174]. Human vocal displays of affective states are determined by internal physiological changes as well as partially determined by social display rules [173]. For example, in formal social situations a person may produce a pleasant voice quality even when internally feeling rage [173]. Research into human voice has identified features of vocal intonation that directly correspond to specific affective states [175, 176]. In [175], it was identified that the mean and range of the fundamental frequency of a verbal utterance can be utilized to distinguish between a specific set of emotions including cold and hot anger, happiness, sadness, anxiety, elation, panic fear, and despair. In [176], a review of empirical studies that have examined the vocal expression of affect during social interactions identified that the intensity (total energy), fundamental frequency, utterance contour, high frequency energy, and word articulation rate can be used to distinguish between the affective states of stress, anger, fear, sadness, joy and boredom. More recent research has also found that the fundamental frequency, spectrum shape, speech rate, and intensity of a voice signal reflect changes in arousal, valence and potency/control during social interactions [177].

For social robots to engage in effective bi-directional communication with humans it is important that they have the ability to perceive and interpret human vocal affect. To-date, a number of affect-recognition systems have been developed for social robots to determine affective states from a person's voice [7, 19, 65, 67, 69, 81, 82, 178, 179]. These systems, similar to the previous systems, can also be classified based on their use by robots in the following HRI scenarios: collaborative [19, 65, 81], assistive [69, 82, 178], mimicry [7], and multi-purpose HRI [67, 179].

5.1 Collaborative HRI

In [19], the role of affect was investigated in HRI scenarios where a PeopleBot robot and a human user jointly completed a task together. The task was to find a signal-transmission location by using verbal communication. The PeopleBot robot utilized the states of the user, itself, and the task to determine which robot actions to perform and which affective states to display. The PeopleBot robot was equipped with two onboard microphones to identify the content and affect of a user's utterance. The affect considered was stress. Unvoiced sounds were removed from the recorded voice signal and a word was determined to be "stressed" if the average frequency of the word had a higher pitch than the current cumulative average pitch. The ratio of stressed words to unstressed words was compared to a threshold to determine if a whole utterance corresponded to a stressed user state. Experiments were conducted with 24 university students. The students controlled the motion of the robot using voice commands such as "turn right" and "go forward" in order for the PeopleBot to locate a signal-transmission location. User stress was induced during the experiments by having the robot warn the user that its battery level was getting low before completion of the activity. The participants were divided into three groups. For the first group, the voice of the robot was modulated to express stress for all battery warnings. For the second group, the voice of the robot only expressed stress in response to the robot detecting that the user was stressed. The third group was a control group and the robot did not express stress. Results indicated that the appropriate expression of affect, as displayed by changes in the robot's voice quality, based on robot state and human affective state improved task performance, namely, that the participants and robot were more successful at finding the signal transmission location.

In [65], the KaMERo robot was designed to engage individuals in games of "20 questions," where the robot asked the questions in order to identify an object the user was thinking of. KaMERo had a mobile base with three articulated legs having wheels and a head with two expressive antenna-like ears anda facial avatar displaying facial expressions. During the game, the robot used an onboard microphone to detect the emotions of happiness, sadness, and anger, from the user's voice in order to determine its own emotional display. The recorded voice signal was used to determine the following features [68]: log energy, Mel-Frequency Cepstral Coefficients (MFCCs), and the corresponding delta and acceleration coefficients. These features were utilized as inputs for a single-state Hidden Markov Models (HMM) trained for each emotion. Experiments with 19 participants, ranging in age from 20 to 40, were conducted where each participant played the game twice with the robot. In one scenario, the robot displayed an emotional behavior (in response to the participant's affective state) and in the next scenario the robot did not display any emotional behavior. The results showed that participants enjoyed the interaction more and found it to be more natural when the robot displayed an emotional behavior. A second set of experiments was performed to validate the emotional-response system of the robot. Ten participants watched videos of the robot playing the game with a person and rated the appropriateness of the robot's responses during the game. The results found that the participants considered the robot's emotional behavior to be appropriate for the interactions.

In [81], a robot arm, Cyton, was used to pick up an object based on the affect of a user. The affective voice of the user was identified using arousal and predictability levels utilizing a voice signal recorded on a microphone headset worn by the user. Affective voice signals were determined by first applying a bandpass filter and computing the energy of the signal. Voiced and unvoiced sounds were identified utilizing an autocorrelation function and the zero-crossing rate applied to the energy of the voice signal. A user's arousal level was identified as the numerical integral on the energy profile of the voiced sound. A user's predictability level was identified as a function of the frequency of occurrence of energy peaks. Experiments with ten university students consisted of the participants speaking to encourage the robot to pick up the object they wanted. The participants were informed that the content of the speech would not influence the robot's behavior, but vocal intonation would. Depending on the affective quality of the utterances, the robot would either hesitate or move with confidence towards an object. The results showed that with vocal affect sensing, the robot was able to pick up the correct object. Post-interaction questionnaires found that although the participants did not have experience with robots they found the system easy to use.

5.2 Assistive HRI

In [82], the small humanoid robot, Nao, acted as a domestic assistant for older adults. A microphone worn on the user was utilized to identify positive and negative valence from a user's voice. A number of features were determined from the recorded voice signal for automated affect classification: fundamental frequency, energy, and MFCCs, relaxation coefficient, phase distortion, the ratio of unvoiced to voiced sound, harmonics-to-noise ratio, jitter, and shimmer. An SVM was utilized to classify these features into positive or negative valence. Experiments consisted of 22 older adults speaking with the Nao robot on topics including well-being, minor illness, depression, medical distress and being happy. Experts were utilized to code the recorded voice signals into a large number of affective states, however to simplify automated classification only voice signals identified as positive or negative valence were investigated. The results of using 2-fold cross validation on the positive and negative valence voice signals of the older adults and investigating voice signal feature performance found the features that best distinguished between negative and positive valence (at a recognition rate between 50 % and 60 %) were phase distortion, shimmer and energy.

In [69] and [178], two automated affect detection systems were designed for future implementation on the T-ROT human-like mobile robot which was developed to assist the elderly. When the automated affect-detection system is implemented, the robot will utilize an onboard microphone to identify the affect of a user speaking to the robot. In [178], the phonemes and pitch of a person's voice were investigated as distinct affective features for use in classification of the affective states of neutral, joy, sadness and anger. First a voice signal was separated into 24 phoneme classes based on the first and second formants of the signal. The average pitch was, then, determined for each phoneme class and each affective state. The results of utilizing this technique were compared to the results of a technique that identified affective state utilizing just the pitch of a voice signal. These techniques were tested on a database of five male and five female actors speaking Korean. The results showed that combining phoneme and pitch information provided probability density functions with smaller standard deviations than the probability density functions of just the pitch information.

In [69], a novel speaker-independent feature was proposed for classifying the affective states of neutral, joy, sadness and anger. The proposed feature was the ratio of spectral flatness to the spectral center of a voice signal. Spectral flatness is the ratio of the geometric mean of the power spectrum to the arithmetic mean of the power spectrum. The spectral center is the average frequency weighted by acoustic power. The proposed approach first identified a voice signal as male or female utilizing pitch and MFCCs features as input to a Gaussian Mixture Model (GMM). Next the pitch, energy and MFCCs were utilized as inputs into another GMM to determine whether the voice signal belonged to one of two groups of affective states: anger and joy, or neutral and sadness. Finally, the proposed ratio of spectral flatness to spectral center was utilized as input for a final GMM to classify the voice signal as one of the affective states in each group. Testing this technique on an updated version of the aforementioned database [178] with the affective states of anger, joy, sadness and neutral, resulted in an average recognition rate of 57.2 %.

5.3 Mimicry HRI

In [7], the Barthoc Jr. child-like robot was used to classify affective voice signals during story telling interactions with users. Using a microphone headset worn by a user, the robot classified a voice signal as communicating happy, fear or neutral affective states. The robot identified the user's affective state using a Naïve Bayes classifier with pitch, MFCCs and energy as input features. An acted voice database [180] was utilized to train the classifier. Experiments consisted of 28 participants reading "Little Red Riding Hood" to the robot while being instructed to generate happy, fear and neutral tones of voice during relevant parts of the story. 17 participants interacted with a robot that would respond to emotional tones of voice by displaying the corresponding facial expression for the detected affective state, and eleven participants interacted with a robot that would respond with a small head nod and a neutral facial expression. Baseline for the affective states was determined by asking the participants to rate the robot on three 5-point Likert scales focusing on whether the robot's facial expression are appropriate for the situation, the robot recognizing the affective components of the story, and the robot's behaviors were similar to that of a human. The results showed that the participants rated the robot that displayed facial expressions higher than the robot that did not display facial expressions.

5.4 Multipurpose HRI

In [179], a cartoon-like entertainment robot was designed to detect a person's affective voice while engaging him/her in a general conversation. The user wore a headset and the affective states of anger, happiness, neutral, sadness and surprise were identified from his/her voice signal. User utterances were extracted by comparing the zero-crossing rate to a predefined threshold. Fourteen fundamental frequency features and energy statistical features were determined from the energy contour of the user's voice signal. These features included the mean, minimum, maximum, median, standard deviation, and derivatives of the fundamental frequency and energy. Fisher's linear discriminant analysis and a hierarchical SVM were utilized to classify the identified features as affective states. Experiments had five participants repeat 15 affective sentences (three for each affective state of neutral, happiness, anger, surprise, and sadness) six times to the robot. The robot displayed a certain behavior in response to each detected affective state, for example, the robot shook its head if it detected that the participant communicated the affective state surprise. The results showed a 73.8 % recognition rate for the five affective states.

In [67], a small round robot, Mung, was developed to detect the affective states of neutral, joy, sadness, and anger during general conversation. An onboard microphone was used to record a user's affective voice when speaking to the robot. A user's utterance was separated into six phoneme groups utilizing the first and second formant frequencies, and the center of gravity of its audio spectrum. Then, six different sets of features were determined and trained for each phoneme group: energy, pitch, zero-crossing rate, log frequency power coefficient, MFCCs and linear prediction coefficient. The mean and standard deviation were calculated for each feature. GMMs were trained for each phoneme group and each set of features. The results of the GMMs were combined to obtain a final user affective state using sum-rule fusion for each set of features. Testing with a database of acted expressions found that the automated affective voice detection system obtained the highest recognition rate of 92.34 % when utilizing MFCCs features. Experiments were, then, performed in which participants were asked to speak to the robot while conveying the affective states of neutral, joy, sadness and anger with their voices. Results of these experiments also found that utilizing only the MFCCs features resulted in the highest recognition rate of 88.33 %.

A number of automated vocal-affect-detection systems have also been proposed for integration and use by robots during social multi-purpose HRI [66, 181–187]. For example, in [181], the Emotional Prosody Speech and Transcripts of the Linguistic Data Consortium database [188] was utilized to test an automated affect detection from voice technique proposed for social robots. The presented technique utilized the zero-crossing rate, log energy, pitch, and MFCCs of a recorded voice signal as input features for GMMs trained for up to 5 affective states. The results found that the technique obtained a recognition rate of 96.9 % when distinguishing between the states of anger and neutral, and 59.4 % when distinguishing between the states of anger, neutral, happy, boredom and sadness.

6 Affective Physiological Signals Recognition During HRI

Human affect influences the body in many ways, for example by changing a person's heart rate, skin conductance and ElectroDermal Response (EDR), tension in specific muscles, breathing rate, etc., [189, 190]. These changes in the body can be monitored as physiological signals and utilized to estimate a person's affective state [189–192]. For example, a decrease in heart rate, measured with an electrocardiogram (ECG) can signify that a person is feeling disgust or sadness. These two affective states can be distinguished from each other by sadness resulting in a decrease in skin conductance, measured with electrodes on the skin, while disgust results in an increase in skin conductance [189]. It has also been found that increased tension in the corrugator muscle (above the eyebrow) and masseter muscle (upper jaw), measured with electromyograms (EMGs), relate to increases in anxiety and mental stress [190]. Additionally, with respect to dimensional models of affect, it has been found that an increase in skin conductance and heart rate relate to an increase in arousal [192]. Physiological signals are well suited for HRI due to data being easily obtained from wireless wearable sensors and analyzed in real-time to detect a user's affective state [193].

Several researchers have developed automated affect-detection systems using physiological signals to allow robots to interpret human affective states during HRI [4, 6, 70, 72, 84–86, 190, 193]. These systems can also be classified based on the proposed HRI scenarios: collaborative [70, 190, 193], assistive [4, 6, 84–86], and multi-purpose [72].

6.1 Collaborative HRI

In [190], a mobile robot was developed for human-robot collaboration during a navigation and exploration task where the human (e.g., an astronaut) and a robot were simultaneously exploring an unknown environment and the robot could respond to the human's affect. The mobile Trilobot robot detected a person's level of anxiety utilizing wearable sensors measuring a user's heart rate, skin conductance, and muscle movement. ECG heart rate data was utilized to calculate parasympathetic and sympathetic nervous system features via power spectral and wavelet packet analysis. Skin conductance features were identified as the mean amplitude and phasic response of the skin conductance signal measured on the fingers. Muscle activity features included the mean amplitude of an electromyogram (EMG) signal recorded over one eyebrow to detect frowning and the variability in the EMG recorded under the jaw to capture clenching/tightening of the jaw. A fuzzy-logic system was utilized to identify a person's anxiety level utilizing the aforementioned features. User self-reported anxiety levels and experimental results were utilized to generate the initial set of rules for fuzzification, which included defining ranges of feature values and generating a set of rules that related these ranges to anxiety levels. An initial set of experiments had participants engaged in mathematical problems with varying difficulty in order to obtain physiological data for varying anxiety levels and training the fuzzy logic system. A second set of experiments had the mobile robot engage in a navigation and exploration task and respond to the physiological data collected from the previous experiment. The robot would detect a user's anxiety level and, then, ask him/her if assistance was needed. If assistance was indeed needed, the robot would trigger an alarm to alert other people that someone needs assistance. The results found a positive correlation between self-reports of a participant's anxiety level and the output of the fuzzy logic system.

In [193], an affect-detection system was designed in order to allow a robot arm to detect a person's affect when the robot and the person are engaged in a collaborative manufacturing or assembly task. An example of such a task is the hand-over-task of an object. The system was designed to detect a person's level of arousal (low, medium, high) and valence (low, medium, high) utilizing physiological data obtained from wearable sensors. HMMs were utilized to classify a person's valence and arousal levels utilizing heart rate, skin conductance, and muscle activity above an eyebrow. An EKG Flex/Pro sensor was utilized to measure heart rate signal in order to determine the peak-to-peak heart rate and heart rate acceleration. A SCFlex-Pro sensor was utilized to measure skin conductance and the rate of change of skin conductance. A Myoscan Pro EMG was utilized to identify the level of response of muscle activity above an eyebrow by applying a low pass filter to the EMG data. All features were utilized as inputs to the HMMs trained for each valence and arousal level. The HMM that resulted in the highest probability for a set of input features was identified as the user's valence or arousal level. Experiments consisted of 36 participants, ages 19 to 56, watching two different robot tasks. The tasks included pick-and-place and reach-and-grab. Each task was generated with two different arm movement planning strategies at 3 different speeds: slow, medium, fast. The planning strategies included a conventional potential field with obstacle avoidance, and safe planning, which in addition to the potential field utilized a danger criterion to minimize the collision force between the robot and human. The participants rated their own levels of arousal and valence for each robot arm movement. Three examples of each valence and arousal level from the experimental physiological data were utilized as training data for the HMMs and the remaining data were utilized for testing. The results showed that the proposed system obtained a recognition rate of 74 % for arousal and 70 % for valence on training data and 61 % for arousal and 62 % for valence on the test data.

In [70], a system was developed to detect a person's affective response to the motion of an industrial robot's arm during HRI. The system detected valence and arousal levels based on wearable sensors which measured facial muscle tension, skin conductance, and heart rate. The same features and sensors described in [193] were used as inputs into a fuzzy inference engine for valence and arousal classification. The input features were fuzzified with trapezoidal membership functions and the rule base was generated using results of physiological research that related each of the features to valence and arousal levels, e.g., when muscle activity above the eyebrow is negative then valence is positive, and when heart rate is faster than baseline then arousal is high. Experiments had ten subjects watch the same set of pick-and-place, and reach-and-grab movements as in [193]. In addition to having their physiological data recorded, the participants rated their own affect on 5-point Likert scales for anxiety, surprise and calm after watching the robot. The results found that the estimated arousal ratings were negatively correlated with subject ratings of calm, and positively correlated with the subjects' ratings of anxiety and surprise. Additionally, subjects reported less anxiety and surprise towards the robot arm motion that utilized the safe planner in comparison to the conventional potential field planner.

6.2 Assistive HRI

In [84], a basketball net was fastened on the end-effector of a 5-degree-of-freedom robot arm in order to allow a person to play a basketball game in which the position and speed of the net was changed to increase or decrease the difficulty of the game. The difficulty of the game was adjusted based on a player's affect and performance level, which was defined as high, medium or low anxiety. The robot utilized the Biopac system to collect physiological data from wearable sensors on the user in order to identify his/her level of anxiety. The identified physiological features included cardiovascular activity via an ECG sensor, EDR via a skin conductance sensor, muscle activity of the cheek, above an eyebrow and on the neck with an EMG as well as skin temperature. These features were utilized as inputs into a SVM for anxiety classification. Experiments were performed in two phases. The first phase consisted of 15 participants engaged in problem solving tasks and playing Pong in order to obtain physiological data to train the SVM for Phase II. Phase II consisted of 14 participants engaged in the basketball game with the robot. The robot arm would vary the difficulty of the game depending on participant performance as well as anxiety level. The results of the experiments showed that the robot was capable of reducing anxiety. Namely, a decrease in anxiety was reported by 79 % of participants, which also resulted in 64 % of the participants improving their performance during the game.

In [6], the above system was expanded to detect the following affective states of a child diagnosed with autism spectrum disorder (ASD) during the basketball game: anxiety, liking, and engagement. The same set of physiological features was used as inputs into SVMs. The SVMs were trained for each investigated affect and ground truth ratings were obtained from a combination of participant, parent, and therapist ratings. Experiments consisted of six participants, ages 13 to 16, playing two basketball games with the robot. The results showed an average recognition rate of 79.5 % for anxiety, 85 % for liking, and 84.3 % for engagement, respectively.

In [4], a mobile nurse robot, PeopleBot, was designed to deliver medicine to elderly residents in a long-term care facility. Users wore a polar heart-rate monitor and an iWorx GSR-200 amplifier used to monitor Galvanic Skin Response (GSR) using electrodes on the index and ring fingers. These physiological signals were used to estimate a user's valence and arousal levels defined as low, medium, and high. In particular, GSR data was processed with wavelet analysis to identify GSR wavelet coefficient features. The mean and standard deviation of the peak-to-peak heart rate were taken as heart rate features. A stepwise backward-elimination regression was applied to these features in order to determine which distinguishable features to use for classification, resulting in 14 GSR wavelet coefficients and the standard deviation of heart rate being used for arousal, and 11 GSR wavelet coefficients, and the mean and standard deviation of the heart rate for valence. These features were utilized as inputs to a NN for affect classification. Experiments consisting of 14 different trials were conducted with 24 residents from two long-term-care facilities. Each subject received a bottle of medicine from the robot during each trial, with each trial having a different physical robot configuration (i.e., varying facial, interaction and voice features). Participants rated their own valence and arousal levels utilizing self-assessment manikins. Baseline physiological data was normalized for each participant prior to the experiments. Utilizing 80 % of the data for training and 20 % for testing, the proposed system obtained a recognition rate of 82 % for arousal and 73 % for valence.

In [85], the OCZ NIA^{TM} headband was utilized to measure muscle tension in a person's forehead in order to estimate his/her stress level. This stress level was then used to control the social behaviors of the Roomba®robot vacuum cleaner. A linear relationship was utilized to determine a user's stress level from the muscle tension measurement. The robot was designed to start vacuuming in a location away from the user if a high level of stress was detected, in an effort to avoid a stressed user. If the robot detected a low level of stress the robot simulated a pet-like behavior by coming closer to the user.

In [86], the mobile robot, PeopleBot, with a face displayed on a screen and a male voice, was utilized as a nurse robot in order to measure a person's blood pressure. The robot would instruct a user how to use an Omron blood-pressure monitor and report the results of the test to the user. Post interaction, the resulting blood pressure was utilized to investigate a person's positive or negative feelings towards the robot. Experiments had 57 participants, all over 40 years old, have their blood pressure measured before and during an interaction with the PeopleBot robot. The participants were also asked to report their affective state during the interaction utilizing the positive and negative affect schedule [194], namely, how they felt towards the robot utilizing 5-point Likert scales on 20 emotional words such as distressed and excited. In addition, they were asked to draw their own ideas for a healthcare robot. The results should that participants who had drawn human-like healthcare robots had higher blood pressure and more negative feelings towards the PeopleBot robot compared to participants who had drawn box-like healthcare robots.

6.3 Multi-Purpose HRI

In [72], a mobile non-expressive human-like robot, PR2, and the expressive human-like mobile dexterous social robot, MDS, engaged users in drawing tasks. Near-InfraRed Spectroscopy (NIRS) brain imaging was employed in order to determine if a person was feeling aversion or affinity towards a robot. The NIRS data was obtained from a two-channel NIRS oximeter sensor worn on the user's head. The NIRS raw-light attenuation data was converted to hemoglobin concentrations, and noise and drift were removed from the signal using a Savitzky-Golay low pass filter [195] and linear detrending. During HRI, the changes in hemoglobin concentrations were utilized to monitor prefrontal brain activity. Experiments were conducted with 38 participants engaged in the drawing tasks with each robot under three different interaction conditions: 1) a 3$^{\text{rd}}$ person condition in which the participant watched a video of another person performing the drawing task with the robot, 2) a 1$^{\text{st}}$ person view in which the robot interacted with a participant remotely via a video conference on a laptop, and 3) a co-located condition in which the robot was located across a table from the participant. Each participant also rated their experience with each of the robots in the aforementioned conditions in addition to having NIRS data collected. Participants rated their arousal using self-assessment manikins, and the eeriness and human-likeness of the robots using 5-point Likert scales. The results showed that the participants had a significant increase in prefrontal brain activity during the co-located interaction with the MDS robot compared to the PR2 robot, while a significant decrease in prefrontal activity was observed during the 3$^{\text{rd}}$ person condition with the MDS compared to the PR2. No significant changes were observed for the 1$^{\text{st}}$ person interaction condition. Taking into account participant self-reports, an increase in prefrontal activity corresponded with aversion (higher arousal and eeriness ratings) towards the robot. Participants were found to have more aversion towards the expressive MDS robot than the PR2 robot.

Several automated affect detection systems have also been designed to be used for future applications in HRI [71, 87]. For example in [71], an affect detection system with future applications in HRI was developed to identify the affective states of anger, boredom, engagement, frustration and anxiety utilizing heart rate, EDR and muscle activity features. The heart rate features, measured from ECG data, include interbeat interval, relative pulse volume, pulse transit time, heart sound, pre-ejection period. The EDR feature is measured as the skin conductance measured from a sensor on a finger. Muscle activity features are measured with EMGs mounted above the eyebrow, on the cheek, and on the neck. Classification was performed utilizing a variety of learning techniques. Experiments consisted of 15 participants engaging in computer based-cognitive tasks and identify their own affective states during the tasks. The results showed that SVM had the highest recognition rate of 85.8 % for instances not included in the training data, however, regression trees, which obtained a recognition rate of 83.5 %, were ranked the highest in terms of computer memory and time efficiency.

7 Multimodal Affect Recognition During HRI

The systems and techniques discussed above focus on the recognition of one single input mode in order to determine human affect. The use of multimodal inputs over a single input provides two main advantages: when one modality is not available due to disturbances such as occlusion or noise, a multimodal recognition system can estimate affective state using the remaining modalities, and when multiple modalities are available, the complementarity and diversity of information can provide increased robustness and performance [74]. Several researchers have considered the combination of two or more input modes in order to effectively determine human affect during the various HRI applications.

Social robots can interact with people using a number of communication channels in order to establish a social relationship. Human affective states can be inferred from a combination of these communication channels. Multimodal data, however, is more difficult to acquire and process due to the need of many more sensors in order to acquire data from multiple channels, the inherent additional dimensionalities in learning algorithms, and multimodal data fusion [73]. The following section will review various multimodal affect detection systems for robots in HRI settings.

7.1 Assistive HRI

In [73], the multimodal acquisition platform, Human Interaction Pervasive Observation Platform (HIPOP), was implemented in the humanoid robot, Facial Automation for Conveying Emotions (FACE), to determine arousal and pleasure levels of adolescents and young adults with ASD. FACE consisted of a passive body, equipped with a female face that was capable of autonomously producing various human-like expressions, such as happiness, anger, sadness, disgust, fear, and amazement. The study consisted of therapists using the robot in interactions with children with ASD to investigate emotional and social recognition capabilities during treatments. A treatment session consisted of: 1) a familiarization stage, where the therapist introduced FACE and the environment to the subject, 2) imitation stages, where a subject described and imitated the robot's facial expressions and eye gaze direction, and 3) play and spontaneous conversation, where the subject engaged in conversations with FACE. The robot was able to speak through a text-to-speech synthesizer that was manually controlled by a human operator. Multimodal inputs used to determine a child's affect included physiological signals, hand gestures, and human eye gaze direction. Physiological signals were acquired through a sensorized t-shirt [196, 197], which had embedded real-time ECG, accelerometer and respiration sensors. The features that were analyzed were heart rate, heart rate variability and R-R intervals. In addition, physiological signals were also acquired by a glove, which provided EDR, hand gestures and postures signals, and skin conductance signal. These features were collected by electrodes located on the fingers of the glove, and a multi-axial accelerometer around the wrist. Moreover, a wearable eye tracking system equipped with a high-speed camera was used for capturing eye gaze direction. Audio signal of speech was acquired from a wearable microphone, while an environmental microphone recorded environmental sounds. Audio signals were used to extract high-level features that were correlated with the subject's psychophysiological state. Pleasure and arousal levels were extracted using Valenza et al.'s quadratic discriminant classifier [198], where extracted features were processed through a Bayesian Decision Tree in order to provide a confidence percentage for each affect level. According to [198], recognition rates were reported to be greater than 90 % for both arousal and valence in a 40-fold-cross-validation. Furthermore, experiments were conducted to test whether FACE was able to involve subjects with ASD in the interaction with the android without any discomfort. Results showed that some children with ASD could not identify the facial expressions of FACE during imitation, but ECG signals indicated that the children were less stressed than in the familiarization phase of the experiment. During gaze imitation, all children with ASD followed FACE when it moved its attention away. Lastly, 55 % of the children with ASD established a spontaneous conversation with FACE.

7.2 Mimicry HRI

An extension of the approach in [75] was presented in [199], where speech recognition from two microphones was incorporated with the facial expression recognition system in the humanoid robotic head, Muecas. Utilizing both modalities, Muecas was able to imitate human affect (happiness, sadness, anger, neutral, fear) and respond in real-time through the display of facial expressions and head pose. The facial expression recognition system was similar to that in [75] where facial expression recognition was accomplished through Gabor filtering and DBN classification. For speech recognition, the speech rate, pitch, and energy were extracted from the signals received from the audio signal. Speech rate was calculated using Fourier transform, comparing the signal with different speech rates and analyzing the amount of energy at each rate. The speech rate calculated served as an input to the DBN classifier. Results showed a speech recognition rate of 87 % for sadness, 71 % for happiness, 67 % for fear, and 82 % for neutral from 40 candidates. Using both modalities, Muecas showed the ability to imitate the recognized affect through its own facial expressions.

7.3 Multi-Purpose HRI

In [200], a multimodal affect detection system, Robotics Dialog System (RDS), using voice and facial expression information was implemented on the social robot, Maggie. Maggie is a doll-like robot capable of moving its head and arms. The robot was equipped with a video camera and microphone within its body, and a webcam in its mouth. The affective states that were classified were happy, neutral, sad, and surprise. Voice features, such as pitch, flux, energy, signal centroid, signal-to-noise, ratio, and number of words pronounced in a minute, were extracted from audio signals obtained from the robot's onboard microphone. These features were input into an off-line classifier trained using decision tree learning (J48) and decision rule (JRIP) on voice examples from interviews, TV shows, audiobook, and databases with tagged voice corpus. Facial expressions (happy, neutral, sad, surprise) recognition was accomplished through third-party software packages which included the Sophisticated High-Speed Object Recognition Engine (SHORE) for face detection and the Computer Expression Recognition Toolbox (CERT) for affect classification in real-time. CERT uses SVM to classify action units in order to infer facial affect [201]. The two modalities were combined using a decision rule based on calculating the degree of confidence using Bayes Theorem. The preliminary performance of each modality was tested on 40 undergraduate students posing the affective states using voice and facial expressions while standing in front of a mobile Pioneer robot equipped with a 2D camera and an omnidirectional microphone. Results showed that the voice decision tree and decision rule had an average success rate of 56.37 % and 57.10 % respectively, and SHORE and CERT had an average success rate of 75.55 % and 62.66 %, respectively. Using multimodal information, the average success rate of the RDS system was recorded to be 83 %, which is higher than the success rate of each individual modality. In an HRI experiment with the robot Maggie, 16 participants were asked by the robot to express one affective state at a time. Results using both modes showed an overall success rate of 77 % for the RDS system.

In [202], a stuffed animal-like social robot, CuDDler, was developed to recognize and respond to a user's emotional acts from facial and voice modalities. The affective states being detected were neutral, happy, sad, angry, and surprise. The software architecture driving CuDDler consisted of several key independent modules including Facial Expression Recognition (FER), Emotion Sound Event Recognition (ESER), Unified Robotic Framework (URF), and Emotion Expression Engine (EEE). The main task of the FER module was to extract the facial expression status from frontal faces and to recognize a user's affect. Images of the user's face were captured by the robot's webcam. The eye positions of the user were estimated based on the Haar-based eye detector and the detected face was cropped and normalized based on the positions of the two eyes in terms of scaling and in-plane rotation. Local Binary Pattern (LBP) features were extracted from selected patches on the normalized face and a linear SVM classifier was used to determine affect. The main task of the ESER module was to detect and classify emotional sounds, such as crying, laughing, or non-voiced occurrences such as stroking, patching, and punching CuDDler. The acoustic signal was continuously tracked using two microphones and the acoustic signature was extracted. The acoustic signature was, then, used to recognize the sound event. The URF module synchronized FER and ESER information together to achieve an appropriate affective response. The main tasks of the EEE module included the generation of different robot emotions, such as happy, sad, and angry, in response to the detected user affect. For instance, for patting, stroking, and laughing, CuDDler would respond with a happy affective state. Experiments were conducted with CuDDler where more than 100 participants interacted with the robot by touching it and displaying facial expressions. Then, participants were asked to fill a Likert scale questionnaire to indicate whether the robot understood their emotional acts, such as hitting, patting, stroking, and squeezing. According to the 70 questionnaires that were completed, the robot was able to recognize the emotional acts of pat, hit, and stroke and responded appropriately to these situations.

In [90], an affective human-robot communication system for the humanoid robot AMI was developed to communicate bi-directionally with a human. The affect recognition system classified the multimodal data from facial expressions (captured by a CCD camera) and voice (captured by a microphone) into three dominant emotions: happy, sad, and angry. The framework developed allowed the robot to respond to the user's current emotional status. Affect recognition through facial expression, captured by the CCD camera, was accomplished by extracting Ekman's facial expression features for the lips, eyebrows and forehead [130]. Affect recognition through speech was accomplished by extracting phonetic and prosodic features. Each feature vector from the two modalities was trained using a NN with affect classifiers for happy, sad, and angry. The training results of the two modalities were integrated together using decision logic where the final affect was determined by a weighted sum of the results of the two modalities. Moreover, the robot was able to produce its own affect through a synthesizer, which produced facial expressions, dialogue and gestures. The affect synthesizer used a 3D emotion model consisting of the following dimensions: stance (open-close), valence (negative-positive), and arousal (low-high). AMI's affect would always be under open stance since the robot was motivated to be openly involved in interactions with humans. The system was able to recognize human affective status, and determine the appropriate behavior according to the user's current affect. For instance, AMI responded with slow speech when the user's overall multimodal affective state was sadness during an interaction of greeting and consoling.

In [203], the humanoid robot, Nao, was used to simulate a child for adult-child interaction. Interactions consisted of the adult greeting the robot, showing a toy to the child robot, revoking the toy, and saying goodbye. The robot was equipped with four microphones located around its head and a 2D video camera located in between its eyes. The robot was able to infer affect from both human voice and human gait, and mapped this affect to a 4D affect GMM known as SIRE (Speed, Intensity, Irregularity, and Extent). The voice features of speech rate, volume range, high-frequency energy ratio, and pitch range were extracted to determine each of the SIRE parameters, respectively. Mapping to SIRE parameters was accomplished by calculating the Z-score of data points relative to the mean and variance over the dataset and the resulting Z-score was shifted and scaled to a range of [0,1]. Human gait features extracted to determine each of the SIRE parameters included: 1) walking speed, 2) maximum foot acceleration, 3) step timing variance, and 4) maximum step length. The same procedure for voice affect was used to map human gait features to the SIRE affect model. Experiments investigated whether the robot could be trained using affective voice and then be able to recognize affective gait. German utterances from 10 subjects from the EmoDB database [180] were used for training and 28 subjects from the Body-Movement Library [204] were used for testing. Results indicated that the robot was able to recognize affect with a 62 %, 90 %, 43 %, and 55 % accuracy for happiness, sadness, anger, and fear, respectively.

In [205], a robot that recognized affect from facial and vocal expression was proposed for general purpose HRI. The robot was designed on a mobile base with a head equipped with a face screen for projectable facial expressions. Facial and voice affect were categorized into five affective states: neutral, happy, sad, fear, and anger. For recognizing facial affect, the facial image was captured by an onboard 2D camera and facial features (eyebrows, eyes, nose, and mouth) were extracted using PCA, where a feature vector containing detected action units of the face was produced. The feature vector was, then, used to classify facial expressions using a Bayesian Network. The Bayesian Network used the positions of action units to infer a probability for each possible facial expression. Recognizing vocal affect was also done in a similar manner. Voice information was recorded with an onboard microphone and features (sampling frequency, pitch, volume level/energy) were extracted using the Praat toolkit [206]. These features were, then, input to a Bayesian Network to classify the vocal expressions as affect. The Bayesian Network used the sampling frequency, pitch, and energy information to infer a probability associated to each vocal expression. All the probabilities produced by the facial and vocal affect recognizers were, then, input into another Bayesian Network decision tree to ultimately infer the human's affective state. Experiments were performed to test both vocal expression classification and facial expression classification. A total of 450 sentences from 50 audio files were used as training for the vocal expression classifier. Results showed an average classification rate of 80.92 % for testing 129 sentences. For facial expression classification, experiments did not use any databases. Testing was done using a captured video stream from the robot's camera. Classification was considered complete when the Bayesian Network produced a probability of 80 % or higher for any of the affective states.

Other automated multimodal affect recognition systems have also been proposed [74, 207–209], but not yet incorporated in robots engaged in HRI. For instance, in [208], a multimodal affect-recognition system using facial expression and voice information was proposed for multipurpose HRI scenarios. The system recognized the following affective states:

anger, disgust, happiness, fear, neutral, sadness, and surprise. For visual facial expression recognition, an AAM was used to describe and generate both the shape and texture of face from facial features. Face pose and facial features (nose, mouth, and eyes) were recognized by a face-detection module [210], where the module generated a feature-based model for every face found and searched the next frame for similar face model information. The facial features were used to initialize the iterative AAM fitting algorithm. A one-against-all SVM was, then, applied to classify the AAM fitting into the seven affective states. For speech-affect recognition, EmoVoice [211], a framework that contained analysis on emotional speech databases and emotional speech application, was used. Features extracted were based on the prosodic and acoustic properties of speech signals such as pitch, energy, linear regression and range of frequency spectrum of short-term signal segments, length of voiced and unvoiced parts in an utterance, and number of glottal pulses. Furthermore, SVM with a linear kernel was used for classification of the seven affective states. For fusion of both recognition modalities, Bayesian Networks were used where results from the individual facial and voice classifiers were fed into the network and the posteriori probabilities were determined. Performance evaluation of the multimodal recognition system was conducted on the DaFEx database [212]. The overall average recognition rate of the system was 78.17 %, compared to 74.46 % for facial expression recognition, and 61.90 % for voice-affect recognition. Another experiment was conducted on four subjects, where the subjects were asked to display facial expressions of five emotion classes: anger, happiness, neutral, sadness, and surprise with and without speaking. Results showed that the multimodal approach again outperformed the single modal approaches: average emotion recognition rate of 58.15 % for multimodal, 55.39 % for facial expression modality, and 23.63 % for voice modality, respectively.

8 Discussions

A wide variety of affect recognition and classification systems have been reviewed herein for socially interacting robots with respect to automated affect detection of humans during HRI scenarios. The most popular input mode used is facial expressions, while there have been a similar number of systems developed for the other modes as well as multi-modal systems. Categorical models have also been used as the main choice of affect classification from facial expressions [57–60, 75, 102–104, 106–108, 116, 134, 136], body language [62–64], voice [7, 65, 67, 69, 178, 179], and multi-modal inputs [73, 90, 199, 200, 202, 205].

The affect categories noted range from using Ekman's six basic emotions [44] along with a neutral emotion [58, 102, 103, 106, 134], or using smaller subsets of these emotions ranging from three [7, 65, 90, 106–108, 134] to five [67, 75, 104–106, 116, 134–136, 199, 202, 205] affective states. Several of the affect-detection systems designed for social robots have used dimensional models, mainly consisting of valence and arousal scales for recognition from facial expressions [76, 77, 125], body language [5], voice [81, 82], physiological signals [4, 70, 86, 193], and multi-modal inputs [193, 203]. A small number of systems have utilized alternative affect classification scales, such as accessibility [79], engagement [61], predictability [81], stance [90], speed regularity and extent [203], stress [19, 85], anxiety [84, 190], and aversion and affinity [213].

The current systems presented are limited to a small number of categorical affective states or a small number of dimensional levels. However, humans can display a large variety of affective states during social interactions [214] and affect-detection systems that can identify only a small set of states (e.g., two to five) will not allow a robot to effectively participate in natural bi-directional affective communications with humans. Hence, research should continue to investigate a variety of affect categorization models, with future work focusing on expanding the number of categorical states or dimensions explored in order to allow socially interacting robots to effectively respond to the large number of possible affective states/levels displayed by people during social HRI.

Each affect-recognition approach discussed above used sensors and feature extraction techniques unique to each particular modality. The most common method of capturing facial information was using a single 2D camera [57–60, 75, 78, 88, 106, 116, 125, 132, 134–136]. Facial features were mainly extracted from the eyes, eyebrows, lips, and nose region according to FACS [130]. However, recognition approaches and facial feature extraction techniques usually only perform well when the human frontal face is positioned directly in front of the camera. This will not always be the case during HRI. A few research groups have also used stereo vision [102, 104] for better facial affect estimation, but no group, to-date, has used 3D information from 3D sensors in order to infer affect.

The Kinect^{TM} sensor, which provides a system with 2D and/or 3D information, was the most commonly used sensor for identifying affect from body language [5, 63, 64, 79]. A wide range of body language features have been investigated, but the Kinect^{TM} skeleton joint positions were utilized more than any other set of features [63, 64].

For detecting affect from voice, the most common sensor was a microphone worn by a user [7, 81, 82, 179]. The most common features detected included the traditional voice features of pitch and fundamental frequency, which human behavior researchers have directly linked to affect [175], as well as features that have been identified more recently, such as MFCCs.

The majority of systems that detect affect from physiological signals have utilized ECG, EMG and/or skin conductance sensors in order to identify a wide variety of heart rate, facial muscle activity, and skin conductance level measures [4, 6, 70, 84, 85, 190, 193]. One physiological system investigated blood pressure [86] and another investigated prefrontal brain activity measured with a near-infrared spectroscopy [72], however, these systems did not report recognition rates, thus, additional research is required to determine how well these signals can effectively recognize affect.

Multimodal systems contain the most sensors compared to all other modalities. Sensors include ECG and EMG [73, 83] to record physiological signals, microphones [83, 90, 199, 200, 202, 203, 205] to record voice intonations, and 2D cameras to capture facial information [90, 199, 200, 202, 205], and body language information [203]. These sensors have been integrated together in order to extract many features, including heart rate [73, 83], voice pitch [83, 90, 199, 200, 202, 203, 205], gait features [203] and facial features [90, 199, 200, 202, 205]. Future research should continue to investigate a wide range of features for all modes in order to determine which combinations of features result in the highest recognition rates during real-world interactions.

The majority of affect-classification techniques have successfully incorporated learning algorithms in order to identify distinct affective states or levels. For facial-affect detection this has included the use of Binary Decision Trees [58], AdaBoost [102], Multilayer Perceptrons [60], SVMs [103] and SVRs [60, 76], NNs [60, 104–108], and Dynamic Bayesian Networks [75].

Affect-classification techniques using body language have included the use of SVM or KNN [64], nearest neighbor [63], as well as the testing of a variety of learning algorithms [5, 61]. Only a few systems have been developed that incorporate body language features that have been linked to affect in psychology or human behavior studies [5, 79]. Although identifying affective states from body language is still an open research area in psychology and human behavior research [29], a number of findings in these research areas can be applied to automated affective body language detection during social HRI. For example, it has been found that culture and context are important when attempting to identify affect from body language displays [29]. Utilizing such additional information will allow robotics researchers to develop more effective human affective body language detection systems that can recognize natural (non-acted) body language during HRI.

GMMs [67, 69] and SVMs [82, 179] have been the most common learning techniques used for classifying affective voice features during HRI. Naïve Bayes [7] and HMMs [65] have also been investigated. Future research in this area should focus on expanding the types of learning algorithms investigated for affect detection. The affect-detection systems using physiological signals have utilized fuzzy models [70, 190], HMMs [193], SVM [6, 84], and NN [4] learning techniques for classifying affect from identified features. The majority of these systems reported positive correlations between recognized affect and baseline affect, with only three studies reporting recognition rates. Future research on systems for detecting affect from physiological signals should also focus on providing more quantitative results with respect to the recognition of specific affective states/levels as is the standard with the other modes. Classification techniques utilized by multi-modal systems included HMMs [83], Bayesian network [73, 199], SVMs [200, 202], NNs [90], and statistical methods [203].

With respect to the specific HRI scenarios, human-affect recognition has been used in a number of game scenarios with robots. A number of them have been specifically with children, ranging from playing chess [61, 78, 88], playing a quiz game [77], playing a basketball game with children with ASD [6], social and emotional intervention with children with ASD [73], simulating a child to engage in adult-child interactions [203], or playing a movement imitation game [62]. Game scenarios with adults have also included playing 20 questions [65] as well as movement imitation [86]. In these scenarios, the affect-aware robots were mainly providing a collaborative role by playing the game with the user, or they were involved in behavioral mimicry. In [6], the robotic arm moving the basketball hoop was used to investigate the use of robotic therapeutic interventions for children with ASD. Other assistive applications of robots, in particular for the elderly, have also been considered including delivering medicine [4] and food [102], as well as providing assistance with activities of daily living including meal eating [5]. The remaining HRI scenarios have been more general and not application dependent, namely focusing on the ability to detect affect [64, 67, 72, 76, 90, 103, 107, 108, 132, 134–136, 179, 200], or learning to mimic human affect [60, 62, 75, 86, 104–106, 116, 125, 199]. It will be interesting to see these affect recognition techniques integrated and tested in more long-term studies with socially intelligent robots in order for us to move closer to an age where social robots will be integrated into our everyday lives assisting and collaborating us with a wide variety of everyday tasks.

A number of databases have also been used to evaluate affect-recognition methods. Popular databases include the Cohn-Kanade database [113], CMU database [118], JAFFE database [131], DaFEx database [212], EmoVoice [211], and EmoDB database [180]. The majority of existing affect detection systems for social HRI have been tested on acted affect displays from facial expressions [58, 76, 102, 103, 116, 134], body language [62–64], voice [7, 67, 69, 178, 179], and multimodal [75, 199, 200, 203] inputs. In general, systems that rely on physiological signals have mainly concentrated on utilizing non-acted data from real-world interactions [4, 6, 70, 72, 84–86, 190, 193]. A smaller number of single mode affect-detection systems using facial expressions, voice or body language have utilized non-acted data from real-world HRI scenarios [4, 5, 19, 61, 65, 73, 77–79, 81, 82, 88, 102]. Databases and acted evaluations can provide an initial approach to testing the performance of these systems. However, they do not provide the sensory information from the real-world scenarios needed for long-term training or testing of systems being used for affect detection during natural real-world interactions. Experimental studies in intended settings with robots and a large number of participants including the targeted users such as children and the elderly will need to be performed to investigate the performance capabilities of affect-detection systems. Furthermore, investigations into the affect displays of different cultures will need to be conducted to ensure the wide use of such automated affect-detection systems during HRI. Developing systems that are robust to age and cultural backgrounds will allow social robots to engage a larger number of users in natural social HRI.

9 Conclusions

In order for socially interacting robots to be accepted into society, they need to be able to interact naturally with human users. For a robot to naturally interact with a person, it must take into account his/her affective state and respond with its own appropriate behavior. In recent years, a number of automated systems have been developed which allow social robots to detect human affect. These automated affect-detection systems have primarily focused on identifying affect from facial expressions, body language, voice, physiological signals, as well as various combinations of those modalities. They have been proposed for or applied to different HRI tasks, including collaboration, assistance, mimicry, as well as multi-purpose HRI scenarios. This survey paper has presented a discussion on the current state-of-the-art in this area discussing each system with respect to its intended HRI scenario and robot, the sensors and sensory information utilized, the feature extraction and affect classification methods employed, as well as the experiments performed and system performance results. Furthermore, the remaining open challenges that still need to be addressed have been outlined in order to promote the effective development of affect-aware social robots capable of autonomously identifying a person's (people's) affect in real-world human-centered HRI scenarios.

Acknowledgments This work was funded by the Natural Science and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs (CRC) Program.

References

1. Goodrich, M., Schultz, A.: Human-robot interaction: a survey. J. Foundations and Trends in Human-Computer Interaction **1**(3), 203–275 (2007)
2. Valero, A., Randelli, G., Botta, F.: Operator performance in exploration robotics. J. Intell. Robot. Syst. **64**(3-4), 365–385 (2011)
3. Rosenthal, S., Veloso, M.: Is someone in this office available to help me? J. Intell. Robot. Syst. **66**(2), 205–221 (2011)
4. Swangnetr, M., Kaber, D.: Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection. IEEE Trans. Human-Machine Syst. **43**(1), 63–75 (2013)
5. McColl, D., Nejat, G.: Determining the affective body language of older adults during socially assistive HRI. In: Proceedings of the IEEE Int. Conf. on Intell. Robots Syst. pp. 2633–2638 (2014)
6. Liu, C., Conn, K., Sarkar, N., Stone, W.: Online affect detection and robot behavior adaptation for intervention of children with autism. IEEE Trans. Robot. **24**(4), 883–896 (2008)
7. Hegel, F., Spexard, T.: Playing a different imitation game: Interaction with an Empathic Android Robot. In: Proceedings of the IEEE-RAS Int. Conf. Humanoid Robots, pp. 56–61 (2006)
8. Breazeal, C.: Social interactions in HRI: the robot view. IEEE Trans. Syst. Man Cybern. C, Appl. Rev. **34**(2), 181–186 (2004)
9. Keltner, D., Haidt, J.: Social functions of emotions at four levels of analysis. Cogn. Emot. **13**(5), 505–521 (1999)
10. Scherer, K.: Psychological models of emotion. The neuropsychology of emotion, pp. 137–162 (2000)
11. Picard, R.: Affective computing. MIT Press (2000)
12. Sorbello, R., Chella, A., Calí, C.: Telenoid android robot as an embodied perceptual social regulation medium engaging natural human–humanoid interaction. Robot. Auton. Syst. **62**(9), 1329–1341 (2014)
13. Park, H., Howard, A.: Providing tablets as collaborative-task workspace for human-robot interaction. In: Proceedings of the ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 207-208 (2013)
14. McColl, D., Nejat, G.: Meal-time with a socially assistive robot and older adults at a long-term care facility. J. Human-Robot Interaction **2**(1), 152–171 (2013)
15. Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Development and evaluation of interactive humanoid robots. Proc. IEEE **92**(11), 1839–1850 (2004)
16. Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., Nakatsu, R.: Robovie: an interactive humanoid robot. Industrial Robot An Int. J. **28**(6), 498–504 (2001)
17. Hinds, P.J., Roberts, T.L., Jones, H.: Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. J. Human-Computer Interaction **19**(1), 151–181 (2004)
18. Längle, T., Wörn, H.: Human–Robot Cooperation Using Multi-Agent-Systems. J. Intell. Robot. Syst. **32**(2), 143–160 (2001)
19. Scheutz, M., Dame, N., Schermerhorn, P., Kramer, J.: The Utility of Affect Expression in Natural Language Interactions in Joint Human-Robot Tasks. In: Proceedings of ACM SIGCHI/SIGART Conf. Human-Robot Interaction. pp. 226–233 (2006)
20. Feil-Seifer, D., Mataric, M.J.: Socially Assistive Robotics. IEEE Robot. Autom. Mag. **18**(1), 24–31 (2011)
21. Ettelt, E., Furtwängler, R.: Design issues of a semi-autonomous robotic assistant for the health care environment. J. Intell. Robot. Syst. **22**(3-4), 191–209 (1998)
22. Conn, K., Liu, C., Sarkar, N.: Towards affect-sensitive assistive intervention technologies for children with autism. Affective Computing: Focus on Emotion Expression. Synthesis and Recognition, pp. 365–390 (2008)
23. Nejat, G., Ficocelli, M.: Can I be of assistance? The intelligence behind an assistive robot. In: Proceedings of the IEEE Int. Conf. Robotics and Automation, pp. 3564–3569 (2008)
24. Breazeal, C., Scassellati, B.: Robots that imitate humans. Trends Cogn. Sci. **6**(11), 481–487 (2002)
25. Bourgeois, P., Hess, U.: The impact of social context on mimicry. Biol. Psychol. **77**(3), 343–352 (2008)
26. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recogn. **36**(1), 259–275 (2003)
27. Zeng, Z., Pantic, M.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. **31**(1), 39–58 (2009)
28. Calvo, R., D'Mello, S.: Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Trans. Affective Comput. **1**(1), 18–37 (2010)
29. Kleinsmith, A., Bianchi-Berthouze, N.: Affective body expression perception and recognition: A survey. IEEE Trans. Affective Comput. **4**(1), 15–33 (2013)
30. Harish, R., Khan, S., Ali, S., Jain, V.: Human computer interaction-A brief study. Int. J. Managment, IT and Eng. **3**(7), 390–401 (2013)
31. Rajruangrabin, J., Popa, D.: Robot head motion control with an emphasis on realism of neck–eye coordination during object tracking. J. Intell. Robot. Syst. **63**(2), 163–190 (2011)
32. Park, J., Lee, H., Chung, M.: Generation of realistic robot facial expressions for human robot interaction. J. Intell. Robot. Syst. (2014). doi:10.1007/s10846-014-0066-1
33. Powers, A., Kiesler, S.: Comparing a computer agent with a humanoid robot. In: Proceedings of the ACM/IEEE Int. Conf. Human-Robot Interaction, pp. 145–152 (2007)
34. Shinozawa, K., Naya, F., Yamato, J., Kogure, K.: Differences in effect of robot and screen agent recommendations

on human decision-making. Int. J. Human Comput. Stud. **62**(2), 267–279 (2005)

35. Kanda, T., Shiomi, M., Miyashita, Z.: An affective guide robot in a shopping mall. In: Proceedings of the ACM/IEEE Int. Conf. Human-Robot Interaction, pp. 173–180 (2009)
36. Yan, H., Ang Jr., M.H., Poo, A.N.: A Survey on Perception Methods for Human-Robot Interaction in Social Robots. Int. J. Soc. Robot. **6**(1), 85–119 (2014)
37. Martinez, A., Du, S.: A Model of the Perception of Facial Expressions of Emotion by Humans: Research Overview and Perspectives. J. Mach. Learn. Res. **13**(1), 1589–1608 (2012)
38. Niedenthal, P.M., Halberstadt, J.B., Setterlund, M.B.: Being happy and seeing “happy” emotional state mediates visual word recognition. Cogn. Emot. **11**(4), 403–432 (1997)
39. Russell, J.A.: Fernández-Dols, J. M.: The psychology of facial expression (1997)
40. Darwin, C.: The expression of the emotions in man and animals. Amer. J. Med. Sci. **232**(4), 477 (1956)
41. Tomkins, S.: Affect, imagery consciousness, vol. I. The positive affects., Oxford, England (1962)
42. Tomkins, S.: Affect, imagery consciousness, vol. II. The negative affects., Oxford, England (1963)
43. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. **17**(2), 124–129 (1971)
44. Ekman, P., Friesen, W., Ellsworth, P.: Emotion in the human face: Guidelines for research and an integration of findings. Pergamon Press (1972)
45. Bann, E.Y., Bryson, J.J.: The Conceptualisation of Emotion Qualia: Semantic Clustering of Emotional Tweets. Prog. Neural Process. **21**, 249–263 (2012)
46. Barrett, L.F., Gendron, M., Huang, Y.M.: Do discrete emotions exist? Philos. Psychol. **22**(4), 427–437 (2009)
47. Wundt, W.: Outlines of psychology. In: Wilhelm Wundt and the Making of a Scientific Psychology, pp. 179–195 (1980)
48. Schlosberg, H.: Three dimensions of emotion. Psychol. Rev. **61**(2), 81–88 (1954)
49. Trnka, R., Balcar, K., Kuska, M.: Re-constructing Emotional Spaces: From Experience to Regulation. Prague Psychosocial Press (2011)
50. Plutchik, R., Conte, H.: Circumplex models of personality and emotions. Washington, DC (1997)
51. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Curr. Psychol. **14**(4), 261–292 (1996)
52. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
53. Remmington, N.A., Fabrigar, L.R., Visser, P.S.: Reexamining the circumplex model of affect. J. Pers. Soc. Psychol. **79**(2), 286–300 (2000)
54. Rubin, D.C., Talarico, J.M.: A comparison of dimensional models of emotion: evidence from emotions, prototypical events, autobiographical memories, and words. Memory **17**(8), 802–808 (2009)
55. Watson, D., Tellegen, A.: Toward a consensual structure of mood. Psychol. Bull. **98**(2), 219–235 (1985)
56. Barrett, L.F.: Discrete emotions or dimensions? The role of valence focus and arousal focus. Cogn. Emot. **12**(4), 579–599 (1998)
57. Kobayashi, H., Hara, F.: The recognition of basic facial expressions by neural network. In: Proceedings of the IEEE Int. Joint Conf. Neural Networks, pp. 460-466 (1991)
58. Wimmer, M., MacDonald, B.A., Jayamuni, D., Yadav, A.: Facial Expression Recognition for Human-robot Interaction–A Prototype. Robot Vision **4931**, 139–152 (2008)
59. Luo, R.C., Lin, P.H., Wu, Y.C., Huang, C.Y.: Dynamic Face Recognition System in Recognizing Facial Expressions for Service Robotics. In: Proceedings of the IEEE/ASME Int. Conf. on Advanced Intell. Mechatronics, pp. 879–884 (2012)
60. Tscherepanow, M., Hillebrand, M., Hegel, F., Wrede, B., Kummert, F.: Direct imitation of human facial expressions by a user-interface robot. In: Proceedings of the IEEE-RAS Int. Conf. on Humanoid Robots, pp. 154–160 (2009)
61. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: Proceedings of the ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 305–311 (2011)
62. Barakova, E., Lourens, T.: Expressing and interpreting emotional movements in social games with robots. Personal Ubiquitous Comput. **14**(5), 457–467 (2010)
63. Xiao, Y., Zhang, Z., Beck, A., Yuan, J., Thalmann, D.: Human-virtual human interaction by upper body gesture understanding. In: Proceeding of the ACM Symp. on Virtual Reality Software and Technology, pp. 133–142 (2013)
64. Cooney, M., Nishio, S., Ishiguro, H.: Recognizing affection for a touch-based interaction with a humanoid robot. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 1420–1427 (2012)
65. Kim, H.R., Kwon, D.S.: Computational model of emotion generation for human–robot interaction based on the cognitive appraisal theory. J. Intell. Robot. Syst. **60**(2), 263–283 (2010)
66. Lin, Y.Y., Le, Z., Becker, E., Makedon, F.: Acoustical implicit communication in human-robot interaction. In: Proceedings of the Conf. on Pervasive Technologies Related to Assistive Environments, pp. 5 (2010)
67. Hyun, K.H., Kim, E.H., Kwak, Y.K.: Emotional feature extraction method based on the concentration of phoneme influence for human–robot interaction. Adv. Robot. **24**(1-2), 47–67 (2010)
68. Yun, S., Yoo, C.D.: Speech emotion recognition via a max-margin framework incorporating a loss function based on the Watson and Tellegen’s emotion model. In: Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4169–4172 (2009)
69. Kim, E.H., Hyun, K.H., Kim, S.H.: Improved emotion recognition with a novel speaker-independent feature. IEEE/ASME Trans. Mechatron. **14**(3), 317–325 (2009)
70. Kulic, D., Croft, E.: Anxiety detection during human-robot interaction. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 616–621 (2005)

71. Rani, P., Liu, C., Sarkar, N., Vanman, E.: An empirical study of machine learning techniques for affect recognition in human–robot interaction. Pattern Anal. Applicat. **9**(1), 58–69 (2006)
72. Strait, M., Scheutz, M.: Using near infrared spectroscopy to index temporal changes in affect in realistic human–robot interactions. In: Physiological Computing Syst., Special Session on Affect Recogntion from Physiological Data for Social Robots (2014)
73. Lazzeri, N., Mazzei, D., De Rossi, D.: Development and Testing of a Multimodal Acquisition Platform for Human-Robot Interaction Affective Studies. J. Human-Robot Interaction **3**(2), 1–24 (2014)
74. Paleari, M., Chellali, R., Huet, B.: Bimodal emotion recognition. Soc. Robot. **6414**, 305–314 (2010)
75. Cid, F., Prado, J.A., Bustos, P., Nunez, P.: A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using Gabor filtering. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 2188–2193 (2013)
76. Schacter, D., Wang, C., Nejat, G., Benhabib, B.: A two-dimensional facial-affect estimation system for human–robot interaction using facial expression parameters. Adv. Robot. **27**(4), 259–273 (2013)
77. Tielman, M., Neerincx, M., Meyer, J.J., Looije, R.: Adaptive emotional expression in robot-child interaction. In: Proceedings of the ACM/IEEE Int. Conf. on Human-robot interaction, pp. 407–414 (2014)
78. Leite, I., Castellano, G., Pereira, A., Martinho, C., Paiva, A.: Modelling Empathic Behaviour in a Robotic Game Companion for Children?: an Ethnographic Study in Real-World Settings. In: Proceedings of the ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 367–374 (2012)
79. McColl, D., Nejat, G.: Affect detection from body language during social HRI. In: Proceedings of the IEEE Int. Symp. on Robot and Human Interactive Communication, pp. 1013–1018 (2012)
80. Xu, J., Broekens, J., Hindriks, K., Neerincx, M.: Robot mood is contagious: effects of robot body language in the imitation game. In: Proceedings of the Int. Conf. on Autonomous agents and multi-agent Syst., pp. 973–980 (2014)
81. Iengo, S., Origlia, A., Staffa, M., Finzi, A.: Attentional and emotional regulation in human-robot interaction. In: Proceedings of the IEEE Int. Symp. on Robot and Human Interactive Communication, pp. 1135–1140 (2012)
82. Tahon, M.: Usual voice quality features and glottal features for emotional valence detection. In: Proceedings of Speech Prosody, pp. 1–8 (2012)
83. Kulic, D., Croft, E.A.: Affective State Estimation for Human-Robot Interaction. IEEE Trans. Robot. **23**(5), 991–1000 (2007)
84. Rani, P., Liu, C., Sarkar, N.: Affective feedback in closed loop human-robot interaction. In: Proceedings of the ACM SIGCHI/SIGART Conf. on Human-robot interaction, pp. 335–336 (2006)
85. Saulnier, P., Sharlin, E., Greenberg, S.: Using bio-electrical signals to influence the social behaviours of domesticated robots. In: Proceedings of the ACM/IEEE Int. Conf. on Human robot interaction, pp. 263–264 (2009)
86. Broadbent, E., Lee, Y.I., Stafford, R.Q., Kuo, I.H.: Mental schemas of robots as more human-like are associated with higher blood pressure and negative emotions in a human-robot interaction. Int. J. Soc. Robot. **3**(3), 291–297 (2011)
87. Schaaff, K., Schultz, T.: Towards an EEG-based emotion recognizer for humanoid robots. In: Proceedings of the IEEE Int. Symp. on Robot and Human Interactive Communication, pp. 792–796 (2009)
88. Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., Mcowan, P.W.: Multimodal affect modeling and recognition for empathic robot companions. Int. J. Humanoid Robot. **10**(1), 1–23 (2013)
89. Gonsior, B., Sosnowski, S., Buss, M., Wollherr, D., Kuhnlenz, K.: An Emotional Adaption Approach to increase Helpfulness towards a Robot. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 2429–2436 (2012)
90. Jung, H., Seo, Y., Ryoo, M.S., Yang, H.S.: Affective communication system with multimodality for a humanoid robot, AMI. In: Proceedings of the IEEE/RAS Int. Conf. on Humanoid Robots, pp. 690–706 (2004)
91. Lim, A., Ogata, T., Okuno, H.G.: Towards expressive musical robots: a cross-modal framework for emotional gesture, voice and music. EURASIP. J. Audio, Speech, and Music Processing **2012**(1), 1–12 (2012)
92. Keltner, D., Ekman, P., Gonzaga, G.C., Beer, J.: Facial expression of emotion. Handbook of affective sciences, pp. 415–432. Series in affective science (2003)
93. Schiano, D.J., Ehrlich, S.M., Rahardja, K., Sheridan, K.: Face to interface: facial affect in (hu)man and machine. In: Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems, pp. 193–200 (2000)
94. Fridlund, A.J., Ekman, P., Oster, H.: Facial expressions of emotion. Nonverbal Behavior and Communication (2nd ed.), pp. 143-223 (1987)
95. Fridlund, A.J.: Human facial expression: An evolutionary view. Academic Press (1994)
96. Fridlund, A.J.: The new ethology of human facial expressions. In: The psychology of facial expression, pp. 103–127. Cambridge University Press (1997)
97. Yang, Y., Ge, S.S., Lee, T.H., Wang, C.: Facial expression recognition and tracking for intelligent human-robot interaction. J. Intell. Serv. Robot. **1**(2), 143–157 (2008)
98. Tapus, A., Maja, M., Scassellatti, B.: The grand challenges in socially assistive robotics. IEEE Robot. Autom. Mag. **14**(1), 1–7 (2007)
99. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. In: Proceedings of the CVPRW Conf. Computer Vision and Pattern Recognition, vol. 5, pp. 53–53 (2003)
100. Castellano, G., Caridakis, G., Camurri, A., Karpouzis, K., Volpe, G., Kollias, S.: Body gesture and facial expression analysis for automatic affect recognition. Blueprint for affective computing: A sourcebook, pp. 245–255 (2010)
101. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robot. Auton. Syst. **42**(3), 143–166 (2003)

102. Breuer, T., Giorgana Macedo, G.R., Hartanto, R., Hochgeschwender, N., Holz, D., Hegger, F., Jin, Z., Müller, C., Paulus, J., Reckhaus, M., Álvarez Ruiz, J.A., Plöger, P.G., Kraetzschmar, G.K.: Johnny: an autonomous service robot for domestic environments. J. Intell. Robot. Syst. **66**(1-2), 245–272 (2011)
103. Littlewort, G., Bartlett, M.S., Fasel, I., Chenu, J., Kanda, T., Ishiguro, H., Movellan, J.R.: Towards Social Robots: Automatic Evaluation of Human-robot Interaction by Face Detection and Expression Classification. In: Advances in Neural Information Processing Syst., vol. 16, MIT Press (2003)
104. Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L.: Imitation as a communication tool for online facial expression learning and recognition. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 5323–5328 (2010)
105. Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L.: A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. Int. J. Soc. Robot. **6**(4), 633–652 (2014)
106. Kobayashi, H., Hara, F.: Facial Interaction between Animated 3D Face Robot and Human Beings. In: Proceedings of the IEEE Int. Conf. on Syst., Man, and Cybernetics, vol. 4, pp. 3732–3737 (1997)
107. Garcíia Bueno, J., González-Fierro, M., Moreno, L., Balaguer, C.: Facial emotion recognition and adaptative postural reaction by a humanoid based on neural evolution. Int. J. Adv. Comput. Sci. **3**(10), 481–493 (2013)
108. Garcíia Bueno, J., González-Fierro, M., Moreno, L., Balaguer, C.: Facial gesture recognition using active appearance models based on neural evolution. In: Proceedings of the ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 133–134 (2012)
109. Seeing Machines: faceAPI. http://www.seeingmachines.com/product/faceapi/ (2009)
110. Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., McOwan, P.W.: It's all in the game: Towards an affect sensitive and context aware game companion. In: Proceedings of the Affective Computing and Intell. Interaction and Workshops, pp. 1–8 (2009)
111. Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., McOwan, P.W.: Inter-ACT: An affective and contextually rich multimodal video corpus for studying interaction with robots. In: Proceedings of the Int. Conf. on Multimedia, pp. 1031–1034 (2010)
112. Barbaranelli, C., Caprara, G.V., Rabasca, A., Pastorelli, C.: A questionnaire for measuring the Big Five in late childhood. Pers. Individ. Dif. **34**(4), 645–664 (2003)
113. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 46–53 (2000)
114. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Mach. Learn. **37**(3), 297–336 (1999)
115. Wisspeintner, T., van der Zan, T., Iocchi, L., Schiffer, S.: Robocup@ Home: Results in benchmarking domestic service robots. In: RoboCup 2009: Robot Soccer World Cup XIII, pp. 390–401 (2010)
116. Dornaika, F., Raducanu, B.: Efficient Facial Expression Recognition for Human Robot Interaction. In: Computational and Ambient Intelligence, pp. 700–708 (2007)
117. Dornaika, F., Davoine, F.: On appearance based face and facial action tracking. IEEE Trans. Circuits Syst. Video Technol. **16**(9), 1107–1124 (2006)
118. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1615–1618 (2003)
119. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis. John Wiley & Sons (2004)
120. Lee, D.D.: & Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (1999)
121. Fausett, L.: Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Prentice-Hall Int. (1994)
122. Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural Comput. **12**(5), 1207–1245 (2000)
123. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 881–892 (2002)
124. Hara, F.: Artificial emotion of face robot through learning in communicative interactions with human. In: Proceedings of the IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 7–15 (2004)
125. Li, Y., Hashimoto, M.: Effect of Emotional Synchronization using Facial Expression. In: Proceedings of the IEEE Int. Conf. on Robotics and Biomimetics, pp. 2872–2877 (2011)
126. Nagamachi, M.: Kansei engineering: a new ergonomic consumer-oriented technology for product development. Int. J. Ind. Ergon. **15**(1), 3–11 (1995)
127. Viola, P., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
128. Yamada, H.: Models of perceptual judgment of emotion from facial expressions. Japanese Psychol. Rev. (2000)
129. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
130. Ekman, P., Friesen, W.: Facial action coding system. Consulting Psychologists Press Inc. (1977)
131. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding Facial Expressions with Gabor Wavelets. In: Proceedings of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
132. Strupp, S., Schmitz, N., Berns, K.: Visual-based emotion detection for natural man-machine interaction. In: Advanced Artificial Intell. pp. 356–363 (2008)
133. Ekman, P., Friesen, W.: Pictures of facial affect. Consulting Psychologists Press (1976)
134. Luo, R.C., Lin, P.H., Chang, L.W.: Confidence fusion based emotion recognition of multiple persons for human-robot interaction. In: Proceedings of the IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 4590–4595 (2012)

135. Cattinelli, I., Borghese, N.A.: A Simple Model for Human-Robot Emotional Interaction. In: Knowledge-Based Intell. Information and Engineering Syst., pp. 344–352 (2007)
136. Cattinelli, I., Goldwurm, M., Borghese, N.A.: Interacting with an artificial partner: modeling the role of emotional aspects. Biol. Cybern. **99**(6), 473–89 (2008)
137. Ge, S.S., Samani, H.A., Ong, Y.H.J., Hang, C.C.: Active affective facial analysis for human-robot interaction. In: Proceedings of the IEEE Int. Symp. on Robot and Human Interactive Communication, pp. 83–88 (2008)
138. Anderson, K., McOwan, P.W.: A real-time automated system for the recognition of human facial expressions. IEEE Trans. Syst. Man Cybern. Part B **36**(1), 96–105 (2006)
139. Cohn, J.F., Reed, L.I., Ambadar, Z., Moriyama, T.: Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. In: Proceedings of the IEEE Int. Conf. on Syst., Man and Cybernetics, vol. 1, pp. 610–616 (2004)
140. Tsai, C.C., Chen, Y.Z., Liao, C.W.: Interactive emotion recognition using Support Vector Machine for human-robot interaction. In: Proceedings of the IEEE Int. Conf. on Syst., Man and Cybernetics, pp. 407–412 (2009)
141. Hong, J.W., Han, M.J., Song, K.T., Chang, F.Y.: A Fast Learning Algorithm for Robotic Emotion Recognition. In: Proceedings of the Int. Symp. on Computational Intell. in Robotics and Automation, pp. 25–30 (2007)
142. Datcu, D., Rothkrantz, L.J.M.: Facial Expression Recognition with Relevance Vector Machines. In: Proceedings of the IEEE Int. Conf. on Multimedia and Expo, pp. 193–196 (2005)
143. Lee, Y.B., Moon, S.B., Kim, Y.G.: Face and Facial Expression Recognition with an Embedded System for Human-Robot Interaction. Affective Computing and Intell. Interaction **3784**, 271–278 (2005)
144. Shan, C., Gong, S., McOwan, P.: Beyond Facial Expressions: Learning Human Emotion from Body Gestures. In: Proceedings of the British Mach. Vision Conf. pp. 1–10 (2007)
145. Mehrabian, A.: Significance of posture and position in the communication of attitude and status relationships. Psychol. Bull. **71**(5), 359–372 (1969)
146. Montepare, J., Koff, E., Zaitchik, D., Albert, M.: The use of body movements and gestures as cues to emotions in younger and older adults. J. Nonverbal Behav. **23**(2), 133–152 (1999)
147. Wallbott, H.: Bodily expression of emotion. Eur. J. Soc. Psychol. **28**(6), 879–896 (1998)
148. Kleinsmith, A., Bianchi-Berthouze, N.: Recognizing affective dimensions from body posture. Affective Computing and Intell. Interaction **4738**, 48–58 (2007)
149. Gross, M., Crane, E., Fredrickson, B.: Effort-shape and kinematic assessment of bodily expression of emotion during gait. Human Movement Sci. **31**(1), 202–221 (2012)
150. Xu, D., Wu, X., Chen, Y., Xu, Y.: Online dynamic gesture recognition for human robot interaction. J. Intell. Robot. Syst. (2014). doi:10.1007/s10846-014-0039-4
151. Hasanuzzaman, M.: Gesture-based human-robot interaction using a knowledge-based software platform. Industrial Robot An Int. J. **33**(1), 37–49 (2006)
152. Yan, R., Tee, K., Chua, Y.: Gesture Recognition Based on Localist Attractor Networks with Application to Robot Control. In: IEEE Computational Intell. Mag., pp. 64–74 (2012)
153. Suryawanshi, D., Khandelwal, C.: An integrated color and hand gesture recognition control for wireless robot. Int. J. Adv. Eng. Tech. **3**(1), 427–435 (2012)
154. Obaid, M., Kistler, F., Häring, M.: A framework for user-defined body gestures to control a humanoid robot. Int. J. Soc. Robot. **6**(3), 383–396 (2014)
155. Malima, A., Ozgur, E., Çetin, M.: A fast algorithm for vision-based hand gesture recognition for robot control. In: Proceedings of the IEEE Signal Processing and Communication Applic. pp. 1–4 (2006)
156. Waldherr, S., Romero, R., Thrun, S.: A gesture based interface for human-robot interaction. Auton. Robot. **9**(2), 151–173 (2000)
157. Corradini, A., Gross, H.: Camera-based gesture recognition for robot control. In: Proceedings of the IEEE-INNS-ENNS Int. Joint Conf. Neural Networks, vol. 4, pp. 133–138 (2000)
158. Boehme, H.: Neural networks for gesture-based remote control of a mobile robot. In: Proceedings of the IEEE World Congr. on Computer Intell. and IEEE Int. Joint Conf. Neural Networks, vol. 1, pp. 372–377 (1998)
159. Burger, B., Ferrané, I., Lerasle, F.: Multimodal interaction abilities for a robot companion. Comput. Vis. Syst. **5008**, 549–558 (2008)
160. Rogalla, O., Ehrenmann, M.: Using gesture and speech control for commanding a robot assistant. In: Proceedings of the IEEE Int. Workshop Robot and Human Interactive Comm. pp. 454–459 (2002)
161. Becker, M., Kefalea, E., Maël, E.: GripSee: A gesture-controlled robot for object perception and manipulation. Auton. Robot. **6**(2), 203–221 (1999)
162. Gerlich, L., Parsons, B., White, A.: Gesture recognition for control of rehabilitation robots. Cogn. Tech. Work **9**(4), 189–207 (2007)
163. Raheja, J., Shyam, R.: Real-time robotic hand control using hand gestures. In: Proceedings of the Second Int. Conf. Machine Learning and Computing pp. 12–16 (2010)
164. Hall, M., Frank, E., Holmes, G.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter **11**(1), 10–18 (2009)
165. Davis, M., Hadiks, D.: Non-verbal aspects of therapist attunement. J. Clin. Psychol. **50**(3), 393–405 (1994)
166. Ganapathi, V., Plagemann, C.: Real-time human pose tracking from range data. In: Comput. Vision–ECCV 2012, vol. 7577, pp. 738–751 (2012)
167. Microsoft: Kinect for Windows Programming Guide. http://msdn.microsoft.com/en-us/library/hh855348.aspx (2014)
168. Lourens, T., Berkel, R.: Van, Barakova, E.: Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. Robot. Auton. Syst. **58**(12), 1256–1265 (2010)
169. Bianchi-Berthouze, N., Kleinsmith, A.: A categorical approach to affective gesture recognition. Connect. Sci. **15**(4), 259–269 (2003)

170. Samadani, A., Kubica, E., Gorbet, R., Kulić, D.: Perception and generation of affective hand movements. Int. J. Soc. Robot. **5**(1), 35–51 (2013)
171. Glowinski, D., Dael, N., Camurri, A.: Toward a minimal representation of affective gestures. IEEE Trans. Affective Comput. **2**(2), 106–118 (2011)
172. Kim, W., Park, J., Lee, W.: LMA based emotional motion representation using RGB-D camera. In: Proceedings of the ACM/IEEE Int. Conf. on Human-Robot Interaction. pp. 163–164 (2013)
173. Scherer, K.: Expression of emotion in voice and music. J. Voice **9**(3), 235–248 (1995)
174. Johnstone, T.: The effect of emotion on voice production and speech acoustics. University of Western Australia (2001)
175. Scherer, K., Bänziger, T.: Emotional expression in prosody: a review and an agenda for future research. In: Proceedings of the Speech Prosody, pp. 359–366 (2004)
176. Iohnstone, T., Scherer, K.: Vocal communication of emotion. In: Handbook of Emotion, pp. 220–235. Guilford, New York (2000)
177. Goudbeek, M., Scherer, K.: Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. J. Acoust. Soc. **128**, 1322 (2010)
178. Hyun, K., Kim, E., Kwak, Y.: Emotional feature extraction based on phoneme information for speech emotion recognition. In: Proceedings of the IEEE Int. Symp. Robot and Human Interactive Comm. pp. 802–806 (2007)
179. Song, K., Han, M., Wang, S.: Speech signal-based emotion recognition and its application to entertainment robots. J. Chinese Inst. Eng. **37**(1), 14–25 (2014)
180. Burkhardt, F., Paeschke, A., Rolfes, M.: A database of German emotional speech. Interspeech **5**, 1517–1520 (2005)
181. Park, J., Kim, J., Oh, Y.: Feature vector classification based speech emotion recognition for service robots. IEEE Trans. Consum. Electron. **55**(3), 1590–1596 (2009)
182. Hyun, K., Kim, E., Kwak, Y.: Improvement of emotion recognition by Bayesian classifier using non-zero-pitch concept. In: Proceedings of the IEEE Int. Workshop Robot and Human Interactive Comm. pp. 312–316 (2005)
183. Kim, E., Hyun, K., Kwak, Y.: Robust emotion recognition feature, frequency range of meaningful signal. In: Proceedings of the IEEE Int. Workshop Robot and Human Interactive Comm. pp. 667–671 (2005)
184. Kim, E., Hyun, K.: Speech emotion recognition using eigen-fft in clean and noisy environments. In: Proceedings of the IEEE Int. Symp. Robot and Human Interactive Comm. pp. 689–694 (2007)
185. Kim, E., Hyun, K.: Speech emotion recognition separately from voiced and unvoiced sound for emotional interaction robot. In: Proceedings of the Int. Conf. Control, Automation, and Syst. pp. 2014–2019 (2008)
186. Liu, H., Zhang, W.: Mandarin emotion recognition based on multifractal theory towards human-robot interaction. In: Proceedings of the IEEE Int. Conf. Robot. Biomimetics, pp. 593–598 (2013)
187. Roh, Y., Kim, D., Lee, W., Hong, K.: Novel acoustic features for speech emotion recognition. Sci. China Series E: J. Techn. Sci. **52**(7), 1838–1848 (2009)
188. Liberman, M., Davis, K., Grossman, M., Martey, N., Bell, J.: Emotional prosody speech and transcripts. In: Proceedings of the Linguist. Data Consortium, Philadelphia (2002)
189. Kreibig, S.: Autonomic nervous system activity in emotion: A review. Biol. Psychol. **84**(3), 394–421 (2010)
190. Rani, P., Sarkar, N., Smith, C., Kirby, L.: Anxiety detecting robotic system–towards implicit human-robot collaboration. Robotica **22**(1), 85–95 (2004)
191. Smith, C.: Dimensions of appraisal and physiological response in emotion. J. Pers. Soc. Psychol. **56**(3), 339–353 (1989)
192. Fernández, C., Pascual, J., Soler, J.: Physiological responses induced by emotion-eliciting films. Appl. Psychophysiol. Biofeedback **37**(2), 73–79 (2012)
193. Kulic, D., Croft, E.: Affective state estimation for human–robot interaction. IEEE Trans. Robot. **23**(5), 991–1000 (2007)
194. Watson, D., Clark, L., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. **54**(6), 1063–1070 (1988)
195. Savitzky, A., Golay, M.J.: Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**(8), 1627–1639 (1964)
196. Paradiso, R., Loriga, G., Taccini, N.: A wearable health care system based on knitted integrated sensors. IEEE Trans. Inf. Technol. Biomed. **9**(3), 337–344 (2005)
197. Scilingo, E.P., Gemignani, A., Paradiso, R., Taccini, N., Ghelarducci, B., De Rossi, D.: Performance evaluation of sensing fabrics for monitoring physiological and biomechanical variables. IEEE Trans. Inf. Technol. Biomed. **9**(3), 345–352 (2005)
198. Valenza, G., Lanata, A., Scilingo, E.P.: The role of nonlinear dynamics in affective valence and arousal recognition. IEEE Trans. Affective Comput. **3**(2), 237–249 (2012)
199. Cid, F., Moreno, J., Bustos, P., Núñez, P.: Muecas: a multisensor robotic head for affective human robot interaction and imitation. Sensors **14**(5), 7711–7737 (2014)
200. Alonso-Martín, F., Malfaz, M., Sequeira, J., Gorostiza, J.F., Salichs, M.a.: A multimodal emotion detection system during human-robot interaction. Sensors **13**(11), 15549–81 (2013)
201. Littlewort, G., Whitehill, J., Wu, T.F., Butko, N., Ruvolo, P., Movellan, J., Bartlett, M.: The Motion in Emotion—A CERT based approach to the FERA emotion challenge. In: Proceedings of the IEEE Int. Conf. on Automatic Face & Gesture Recognition and Workshops, pp. 897–902 (2011)
202. Limbu, D.K., Anthony, W.C.Y., Adrian, T.H.J., Dung, T.A., Kee, T.Y., Dat, T.H., Alvin, W.H.Y., Terence, N.W.Z., Ridong, J., Jun, L.: Affective social interaction with CuDDler robot. In: Proceedings of the IEEE Int. Conf. on Robotics, Automation and Mechatronics, pp. 179–184 (2013)

203. Lim, A., Member, S., Okuno, H.G.: The MEI Robot?: Towards Using Motherese to Develop Multimodal Emotional Intelligence. IEEE Trans. Auton. Ment. Dev. **6**(2), 126–138 (2014)
204. Ma, Y., Paterson, H., Pollick, F.: A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behav. Res. Methods **38**(1), 134–141 (2006)
205. Prado, J.A., Simplício, C., Lori, N.F., Dias, J.: Visuo-auditory multimodal emotional structure to improve human-robot-interaction. Int. J. Soc. Robot. **4**(1), 29–51 (2011)
206. Boersma, P.: Praat, a system for doing phonetics by computer. Glot Int. **5**(9-10), 341–345 (2001)
207. Kwon, D., Kwak, Y.K., Park, J.C., Chung, M.J., Jee, E., Park, K., Kim, H., Kim, Y., Park, J., Kim, E., Hyun, K.H., Min, H., Lee, H.S., Park, J.W., Jo, S.H., Park, S., Lee, K.: Emotion interaction system for a service robot. In: Proceedings of the IEEE Int. Symp. on Robot and Human interactive Communication, pp. 351–356 (2007)
208. Rabie, A., Handmann, U.: Fusion of audio-and visual cues for real-life emotional human robot interaction. In: Pattern Recognition, vol. 6835, pp. 346–355 (2011)
209. Yoshitomi, Y., Kim, S.I., Kawano, T., Kilazoe, T.: Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face. In: Proceedings of the IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 178–183 (2000)
210. Castrillón, M., Déniz, O., Guerra, C., Hernández, M.: ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. J. Vis. Commun. Image Represent. **18**(2), 130–140 (2007)
211. Vogt, T., André, E., Bee, N.: EmoVoice—A framework for online recognition of emotions from voice. In: Perception in Multimodal Dialogue Syst., pp. 188–199 (2008)
212. Battocchi, A., Pianesi, F., Goren-Bar, D.: A first evaluation study of a database of kinetic facial expressions (dafex). In: Proceedings of the Int. Conf. on Multimodal Interfaces, pp. 214–221 (2005)
213. Strait, M., Scheutz, M.: Measuring users' responses to humans, robots, and human-like robots with functional near infrared spectroscopy. In: Proceedings of the IEEE Int. Symp. Robot and Human Interactive Communication, pp. 1128–1133 (2014)
214. Hareli, S., Parkinson, B.: What's social about social emotions. J. Theory Soc. Behav. **38**(2), 131–156 (2008)

Derek McColl received his B.Sc. (Eng) degree in mathematics and engineering and M.A.Sc. degree in mechanical engineering from Queen's University, Kingston, ON, Canada, in 2007 and 2010, respectively. He received his Ph.D. in mechanical engineering from the University of Toronto, Toronto, Canada, in 2015. He is currently a post-doctoral fellow at Defence Research and Development Canada. His research interests include robotics, mechatronics, human–robot interaction and affect recognition.

Alexander Hong received his B.A.Sc. degree in Engineering Science (Aerospace) from the University of Toronto, Toronto, Canada, in 2014, where he is currently pursuing his M.A.Sc. degree in mechanical engineering under the supervision of Dr. Benhabib and Dr. Nejat, with a focus on affect recognition in healthcare robots.

Naoaki Hatakeyama received his B.Eng. degree in mechanical engineering from Meiji University, Kanagawa, Japan, in 2014. He is currently pursuing his M.Sc. degree at Chiba Institute of Technology, Chiba, Japan. His research interests include robotics, and human-robot interaction.

Goldie Nejat received her B.A.Sc. and Ph.D. degrees in mechanical engineering from the University of Toronto, Toronto, Canada, in 2001 and 2005, respectively. She is currently an Associate Professor and the Director of the Autonomous Systems and Biomechatronics Laboratory (ASBLab) in the Department of Mechanical and Industrial Engineering at the University of Toronto. She is also the Director of the Institute for Robotics and Mechatronics (IRM) at the University of Toronto, and an Adjunct Scientist at the Toronto Rehabilitation Institute, Toronto, ON, Canada. Her current research interests include affect sensing, human–robot interaction, semi-autonomous and autonomous control, and intelligence of assistive/service robots for search and rescue, exploration, healthcare, and surveillance applications.

Beno Benhabib received his B.Sc., M.Sc., and Ph.D. degrees all in mechanical engineering in 1980, 1982, and 1985, respectively. He has been a Professor with the Department of Mechanical and Industrial Engineering, and the Institute for Biomaterials and Biomedical Engineering at the University of Toronto, Toronto, Canada, since 1986. His current research interests include the design and control of intelligent autonomous systems.