J Multimodal User Interfaces (2010) 3: 89–98
DOI 10.1007/s12193-009-0033-5

ORIGINAL PAPER

# Affect recognition for interactive companions: challenges and design in real world scenarios

**Ginevra Castellano · Iolanda Leite · André Pereira · Carlos Martinho · Ana Paiva · Peter W. McOwan**

Received: 16 April 2009 / Accepted: 25 November 2009 / Published online: 16 December 2009

**Abstract** Affect sensitivity is an important requirement for artificial companions to be capable of engaging in social interaction with human users. This paper provides a general overview of some of the issues arising from the design of an affect recognition framework for artificial companions. Limitations and challenges are discussed with respect to other capabilities of companions and a real world scenario where an iCat robot plays chess with children is presented. In this scenario, affective states that a robot companion should be able to recognise are identified and the non-verbal behaviours that are affected by the occurrence of these states in the children are investigated. The experimental results aim to provide the foundation for the design of an affect recognition system for a game companion: in this interaction scenario children tend to look at the iCat and smile more when they experience a positive feeling and they are engaged with the iCat.

G. Castellano (✉) · P.W. McOwan
Department of Computer Science, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
e-mail: ginevra@dcs.qmul.ac.uk

P.W. McOwan
e-mail: pmco@dcs.qmul.ac.uk

I. Leite · A. Pereira · C. Martinho · A. Paiva
INESC-ID, Instituto Superior Técnico, Porto Salvo, Portugal

I. Leite
e-mail: iolanda.leite@tagus.ist.utl.pt

A. Pereira
e-mail: andre.pereira@tagus.ist.utl.pt

C. Martinho
e-mail: carlos.martinho@tagus.ist.utl.pt

A. Paiva
e-mail: ana.paiva@inesc-id.pt

**Keywords** Interactive companions · Affective and social robotics · Affect recognition · Affective cues · Socially intelligent behaviour

## 1 Introduction

Building artificial companions capable of interacting with humans in the same way that humans communicate with each other has always been a major goal of the research in artificial intelligence. Current prototypes of artificial companions (e.g., robots, virtual agents, etc.) are still lacking important capabilities, which often impedes human users to establish bonds with them.

One of the main requirements for an artificial companion to be able to engage in a natural interaction with human users is the ability to display social, affective behaviour [1, 2]. Social capabilities are necessary for all those applications in which a robot or a virtual agent needs to interact with humans as a companion, a partner or a friend [3]. Companions can represent a valuable resource in many different contexts. For example, they can offer assistance in the care for elderly or in therapy applications, with the purpose not to replace human contact but to provide additional functionalities. Companions can be used as personal assistants for domestic applications or to help the user wherever they go (e.g., at work, on a trip, etc.). They can also be employed to entertain and motivate users, for example for edutainment purposes or in entertainment industry (e.g., socially interactive games).

Interactive companions must be capable of sensing, processing and interpreting information about the user and the context in which the interaction takes place, in order to be able to plan and generate an appropriate response. Understanding the user's affective or mental states from their verbal and more subtle non-verbal behaviour is of vital importance for a companion to be able to act socially. A socially intelligent companion, for example, would try to ensure that the user is interested in maintaining the interaction or to act empathically towards them if they are sad or not willing to engage in an interaction, e.g., it would not disturb them trying to engage them in some activity if they do not approach it.

The ability to attribute affective or mental states to the user can be referred to as "affect sensitivity". It refers to the way social affective cues conveyed by people's behaviour can be used to infer behavioural states, such as affective or mental states. These span from basic emotions (such as joy, anger, sadness, etc.) to more complex affective and mental states such as interest, boredom, frustration, etc.

This paper reviews some of the issues arising in the process of endowing interactive companions with affect sensitivity. Challenges in the design of an affective perceptual framework able to work over extended periods of time are discussed and issues related to the need for systems to work in real-time and real environments are examined. Finally, a naturalistic scenario in which an iCat robot plays chess with children is presented. In this scenario, the role of non-verbal behaviour is investigated for the modelling of user states related both to the task and the social interaction with the robot. Feeling and level of engagement with the iCat experienced by the children are selected as the states that the iCat robot should be sensitive to in order to be able to sustain long-term interactions with users and the non-verbal behaviours affected by the occurrence of these states in the children are identified. Statistical analysis shows that children tend to look at the iCat and smile more in correspondence with positive feeling and engagement with the iCat. Furthermore, additional information about the phase of the game proved helpful in determinating when children look at the iCat more in correspondence with a specific type of feeling and level of engagement. We aim to exploit these results to carry out a rigorous design of an affect recognition system for a game companion based on the analysis of multimodal non-verbal behaviour.

## 2 Affect recognition for interactive companions

Endowing a robot or a virtual agent with affect sensitivity is a complex task. Researchers have been increasingly addressing the design of systems endowed with this ability. Nevertheless, the attempts that have been made towards the development of virtual agents and robots capable of inferring the user's states in real-time are still not numerous. Further, not many system prototypes have been designed which can work in real environments in the long term. The need for artificial companions to work in the user's own social settings and to create long-term relationships with humans requires then research on affect recognition to be taken beyond the state of the art. In the following we review some of the challenges and relevant issues for the design of interactive companions from the perspective of affect recognition.

### 2.1 Beyond prototypical emotions

Many of the affect recognition systems described in the literature mainly focused on the recognition of basic emotions (e.g., joy, sadness, disgust, surprise, fear, anger, etc.) [4]. While the automated recognition of more complex states has started to receive some attention only lately [4], research on artificial companions requires the design of an affective framework in which the companion's affect sensitivity goes beyond the ability to recognise prototypical emotions, and allows for more variegated affective signals conveying more subtle states such as, for example, boredom, interest, frustration, agreement, etc., to be captured. To date, some efforts have been made to detect such non basic affective states. Kapoor et al. [5], for example, presented an automated method to predict frustration using multimodal non-verbal cues including facial expressions, head movement, posture, skin conductance and mouse pressure data. El Kaliouby and Robinson [6] developed a computational model that detects in real-time complex mental states such as agreeing, concentrating, disagreeing, interested, thinking and unsure from head movement and facial expressions in video. Yeasin et al. [7] proposed an approach that recognises six universal facial expressions and uses them to compute levels of interest. Peters et al. [8] modelled user level of interest and engagement using eye gaze and head direction information during an interaction with a virtual agent displaying shared attention behaviour. Castellano et al. [9] proposed a framework for the modelling of engagement with a robot companion using task and social interaction-based features.

Other attempts to infer other affective expression nuances have been reported in the literature. Caridakis et al. [10] presented a dynamic approach based on facial expressions and speech data to interpret coarse affective expressions in terms of a dimensional space representing activation and valence. Castellano et al. [11] proposed an approach to analyse emotional expression in music performance aiming to discriminate emotionally expressive intentions (i.e., sad, allegro, serene, personal and overexpressive) of a pianist using movement expressivity features. Some efforts have been made to detect other complex mental states such as shame, depression or pain. Littlewort et al. [12], for example, used an automated system for facial expression recognition to detect expressions of pain. See Zeng et al. [4] for an extensive overview.

It is important to stress that the inclusion of affect representation into a framework for affect recognition is of primary importance. Incorporating models and paradigms developed by psychologists for the classification of affective states [13] is a pressing need, but is still a challenging issue. Strengthening the connection with psychological models would allow for the first steps towards the detection of more complex affective states (e.g., appraisals, blends of emotions, preferences, mood, attitudes, etc.) to be undertaken.

### 2.2 Spontaneous versus acted

The design of most existing affect recognition systems was largely based on databases of acted affective expressions [4]. While acted affective expressions, contrary to spontaneous expressions, can be defined precisely, allow for the recording of several affective expressions for the same individual, and can be characterised by very high quality, they often reflect stereotypes and exaggerated expressions, not genuine affective states, and they are often decontextualised [14].

To date, some efforts to develop systems for the automatic analysis and detection of spontaneous affective expressions have been reported. Examples include the neurofuzzy system for emotion recognition by Ioannou et al. [15], which allows for the learning and adaptation to specific users' naturalistic facial expressions, the approach proposed by Caridakis et al. [10], which models affective expressions from naturalistic video sequences using facial expressions and speech prosody-related features, the system by Kapoor et al. [5], which uses multimodal non-verbal cues to detect frustration in students using a learning companion, the work by Devillers and Vasilescu [16], who reported results on detection of affective states in a corpus of real-life dialogs collected in call centers using linguistic and paralinguistic features. Noteworthy efforts are those of Valstar et al. [17] and Littlewort et al. [12], who reported results on the automatic discrimination between posed and spontaneous facial expressions.

The design of an artificial companion would certainly benefit from the development of affect detectors which are trained and tested with spontaneous, real-life expressions. Collection of naturalistic data involves several issues, such as the difficulty of recording several emotional reactions for the same individual. Nevertheless, this is an issue that must be addressed in the design of an affect sensitive companion, in which personalisation plays an important role.

### 2.3 Multimodal affective expressions

Another important issue for affect sensitive artificial companions is the need for a multimodal affect recognition system. It is expected that a companion is endowed with the ability to analyse different types of affective expressions, depending on the specific scenario of interaction with the user. On the other hand, fusing different affective cues can allow for a better understanding of the affective message communicated by the user to be achieved. While unimodal systems (mainly based on facial expression or speech analysis) have been deeply investigated, studies taking into account the multimodal nature of the affective communication process are still not numerous [4]. To date, some efforts to combine two modalities of expressions for the purpose of affect recognition have been reported in the literature, namely based on the fusion of facial expressions and body gesture data (e.g., [18]), facial expressions and head gesture (e.g., [6]), head and body gesture (e.g., [11]), facial expressions and speech (e.g., [10]), physiological signals and speech (e.g., [19]). Some attempts of using multiple modalities include the system developed by Kapoor et al. [5], which allows for the detection of frustration using multimodal non-verbal cues such as facial expressions, blink, head movement, posture, skin conductance and mouse pressure, the work by Castellano et al. [20], in which facial expressions, body gesture and speech data is fused at the feature and decision level to predict eight emotional states in a speech-based interaction, the study conducted by Valstar et al. [17], which combines multimodal information conveyed by facial expressions, head and shoulders movement to discriminate between posed and spontaneous smiles.

An issue of primary importance in multimodal affect recognition is represented by the fusion of different modalities. Features from different modalities of expressions can be fused at different levels (e.g., feature-level or decision-level fusion). Results from studies in psychology and neurology [21, 22] suggest that the integration of different perceptual signals occur at an early stage of human processing of stimuli. This suggests that different modalities of expression should be processed in a joint feature space rather than combined with a late fusion. Features from different modalities are often complementary and redundant and their relationship is unknown. For this reason it is important to highlight that combination schemes other than the direct feature fusion must be investigated. The development of novel methods for multimodal fusion should take into consideration what are the underlying relationships and correlation between the feature sets in different modalities [23, 24], how different affective expressions influence to each other and how much information each of them provides about the communicated affect.

### 2.4 Working in real environments

Artificial companions have to be designed so as to be able to work in the users' own settings. This requires a companion's affect recognition system to be robust in real world conditions: face detectors and body and facial features tracking systems which are robust to occlusions, noisy background (e.g., shadows, illumination changes) [25], rigid head motions [26], etc., are some of the most important requirements for a companion to successfully work in real environments.

Real world scenarios means that the companion must infer the user's state in real-time. This poses several issues, such as, for example, the segmentation and the analysis of the temporal dynamics of face or body gestures and expressions, since the possibility for a user's affective state to start at any time is a crucial factor in real-time affect recognition [4]. Further, detecting blends of affective states is still an open issue, as well as establishing their boundaries.

The dynamics of affective expressions is a primary factor in the understanding of human behaviour. Affective expressions vary over time, together with their underlying affective content: analysis of static affect displays cannot account for the temporal changes in the response patterning characterising an emotional reaction. It is therefore crucial for artificial companions to be capable of analysing the temporal dynamics of affective expressions and their temporal correlation in order to be able to establish a long-term interaction with the user, which is dynamic by definition. Detection of temporal segments of affective expressions and analysis of their temporal evolution are then primary issues in the design of an affect recognition system for artificial companions. To date, some efforts towards a dynamic account for affective expressions have been reported in the literature. Examples include the work by Pantic and Patras [25], which deals with facial actions dynamics recognition, the study by Valstar et al. [17], which showed the important role of the temporal dynamics of face, head and shoulder expressions in discriminating posed from spontaneous smiles, and the works by Castellano et al. [11, 27], which investigated the role of the dynamics of the expressivity of human movement (specifically upper-body and head gestures) and reported that features related to the timing of movement expressivity (e.g., expressive motion cues such as the quantity of motion and the velocity) are effective for the discrimination of affective states.

## 2.5 Personalisation

Individual differences cannot be neglected while designing an affect recognition system for artificial companions. People differ for culture, gender, personality, preferences, goals, etc. and do not express the same affective state in the same way. Affect sensitive artificial companions must then be endowed with a personalised affective framework: their recognition abilities must be designed so as to be person-dependent and to adapt to a specific user over time. Ioannou et al. [15], for example, developed a neurofuzzy rule-based system in which an initial set of rules can be modified via a learning procedure to adapt to a specific user's affective facial expressions.

An important issue to be considered in the design of an affect recognition system for artificial companions is represented by taking into consideration the context in which an affective expression is displayed (e.g., characteristics of the person expressing the emotion, environment in which the emotion is displayed, what the person is doing (i.e., their task), underlying mood, behaviour displayed by the companion, presence of other people, etc.). As suggested by Scherer [28], there can be as many emotions as the patterns of appraisal results. This highlights the importance of the evaluation of a stimulus event for the generation of the emotional response. In the same way, artificial companions must be able to evaluate how the recognised affective state relates with the conditions external to an individual that elicited the emotional response. See [9, 29] for an example of systems in which the role of context in affect recognition is addressed.

Finally, artificial companions must be designed so as to be capable of engaging in a long-term relationship with users (e.g., over periods of weeks or months). Previous studies showed that the novelty effect often quickly wears out [30] and people change their attitude towards the companion over time and their engagement decreases. For the purpose of the design of an affect recognition system for artificial companions this is an issue that must be addressed. An affect recognition framework which is adaptive over time (e.g., designed so as to work with rules that change according to how the level of engagement of the user towards the companion varies over different sessions of interaction) is required.

## 2.6 Closing the affective loop

Affect sensitivity is a prerequisite for a companion to act socially and generate responses appropriate to the user's behaviour, thus it affects other capabilities of companions.

An artificial companion must be endowed with mechanisms for expressing social, affective behaviour. These should include non-verbal (e.g., facial and bodily expressions of the companion) and verbal communicative expressions. These mechanisms should be integrated with an architecture that includes memory, emotion, personality, adaptation and autonomous action-selection (see, for example, [31]). Adaptive models and mechanisms that support both collaborative and autonomous decision-making influenced by the companion's internal state and past experiences must be developed.

Information on the user's affective states can be used to modulate the companion's internal state, memory, expressive and cognitive behaviour such as decision-making and planning. For example, a companion needs to know what to remember and what to forget. Hence, memory of companions can be designed so that emotional episodes can be remembered more and information about the user's affective state during specific sessions of interaction can be used to retrieve relevant information when new interactions take place.

Moreover, analysis of the user's affective behaviour can be used to influence the way a companion acts or communicates. For example, a companion can respond to low-level affective cues such as the way the user gestures (e.g., the expressivity of their movement) or to higher level data abstractions such as recognised facial expressions by exhibiting a low-level generated affective behaviour (e.g., affective copying, mimicry) [32], or try to infer the user's affective state in order to plan a more complex response.

## 3 Design in real world scenarios: a case study

This section provides an example of a real world scenario for an affect sensitive artificial companion: an iCat robot that plays chess with children. Challenges from the perspective of affect recognition are discussed in this interaction scenario.

Researchers working in the field of intelligent tutoring systems aim to develop agents that act as peer tutors capable of detecting the user's affective state, as it is acknowledged that students become more motivated when the tutor provides affective support and empathy towards the learner [33]. Arroyo et al. [34], for example, proposed a model for affect detection based on input from multiple sensors, the learner's own feedback of their emotional state and the context of the performed task. An experiment performed in a public school environment showed how most of the affective states that the users experienced were related to the task they performed.

Similarly, we show how, in our scenario, it is necessary the companion to be sensitive to application-dependent affective states, which are very different from prototypical emotions. Multimodal non-verbal behaviour is analysed in relation to the identified states. Statistical analysis was performed to identify which behaviours allow for a discrimination of the identified affective states in this interaction scenario.

### 3.1 Towards an affect sensitive game companion

This scenario involves a game companion for young students that plays educational games such as chess. Chess helps children develop their memory, logic and problem-solving skills [35, 36]. In general, children that receive systematic chess instructions are more likely to improve their school efficiency in different subjects [37].

The scenario includes a social robot, the iCat [38], which plays chess with children using an electronic chessboard. The iCat helps children improve their chess skills: while playing with the iCat, they receive feedback about their moves on the chessboard through the robot's facial expressions, which are generated by an affective system influenced by the state of the game (for more details on the affective system and on how the iCat's facial expressions are generated, see [39]). Note that the children do not know beforehand that the facial expressions displayed by the robot are supposed to help them improve their skills. The affective system is self-oriented, i.e., when the user makes a good move the iCat displays a sad facial expression and when the user makes a bad move the iCat displays positive reactions. We have adopted this approach instead of a more cooperative behaviour because, from observations of children playing against each other during chess lessons, these reactions were more consistent with what they expected about their opponents.

The interaction starts with the iCat waking up and inviting the user to play. After each move made by the user, the robot generates an affective reaction. Afterwards, the iCat asks the user to make its move, as its embodiment does not allow it to do so itself. When the user makes the move that the iCat requested, the iCat sends a confirmation signal to the user, for example a small utterance such as "ok, thank you" or a nodding). The game continues until one of the opponents checkmates the other.

This interaction scenario poses several challenges from the point of view of affect recognition. Off-line analysis of videos recorded during several interactions showed that children display prototypical emotional expressions only occasionally. For this reason, we believe that the interaction in our scenario may benefit from the companion's assessment of other affective nuances of the user, for example the user's feeling (e.g., positive vs. negative) and engagement with the iCat. The *valence of the feeling* experienced by the user [40] was selected to describe the overall feeling that the user experiences during the interaction. The companion may use non-verbal affective cues displayed by the user (e.g., cues typical of a face-to-face interaction, such as facial expressions, eye gaze, head movements, body pose, etc.) to infer the user's affective or mental state and plan and change its actions during the game in an appropriate way. The companion's sensitivity towards social, affective cues of this kind may contribute to underpin an empathic interaction. The analysis of the user's behavioural cues may also be exploited for the purpose of monitoring the user's level of engagement with the iCat over time, within each session and over different sessions of interaction. This is especially important to evaluate how the acceptance and user experience vary in the long term and requires the design of a dynamic, adaptive affective framework. Hence, the user's *engagement with the iCat* was chosen to describe the level of social interaction established between them, i.e., the user's willingness to start and maintain the interaction with the iCat [41].

### 3.2 Data collection

In order to identify what type of non-verbal behaviours displayed by the users when interacting with the iCat are more affected by the identified affective states (i.e., *feeling* and *engagement with the iCat*), an experiment was conducted at a primary school where children have chess lessons as part of their extracurricular activities. Children were used to learn to play chess, as they played two hours per week supervised by a chess instructor for more than one year.

5 eight year old children (3 male and 2 female) took part in the experiment, playing two different exercises with the iCat. The exercises consisted of middle game chess positions suggested by the school's chess instructor, who was familiar with each student's chess skills. In the attempt to induce different types of affective states and behaviours in the participants, the two exercises were selected so as to be different in terms of difficulty: one of them was characterised by a low level of difficulty and the second one by a medium level. The duration of each exercise varied depending on the specific participant, with exercises lasting up to 15 minutes at most.

All the interactions were recorded with three video cameras: one capturing the frontal view (e.g., the face of the children, see Fig. 1), one the side view and one the iCat. The videos recorded with the frontal camera were annotated in terms of user states by three trained annotators (two researchers in computer science and one in psychology) who worked separately. The annotation was based on the behaviour displayed by the children and the situation of the game.

72 video segments were randomly selected starting from the collected videos using ANVIL, a free video annotation tool [42], and a predefined scheme was created for annotation. Each video segment had a duration of 7 seconds at most. Before starting the annotations, the annotators agreed on the meaning of each label to describe the user state in the videos. As for the valence of the feeling, annotators could choose one out of three options: "positive", "negative" and "cannot say". To describe the engagement with the iCat, annotators could choose among "engaged with the iCat", "not engaged with the iCat" and "cannot say". As the off-line analysis of videos with the recorded interactions showed that the participants do not display prototypical or exaggerate expressions, it was decided not to include any "neutral" class in order to highlight the differences between extreme conditions of the same user state. The three annotations were compared for each video segment: a label was selected to describe the state of the user in a video segment when it was chosen by two or three of the annotators. In case each of the annotators chose a different label, the video segment was labelled as "cannot say". From the annotation process, we randomly selected 15 video segments labelled as "positive", 15 as "negative", 15 as "engaged with the iCat" and 15 as "not engaged with the iCat". Each group of videos contains 3 samples for each participant.

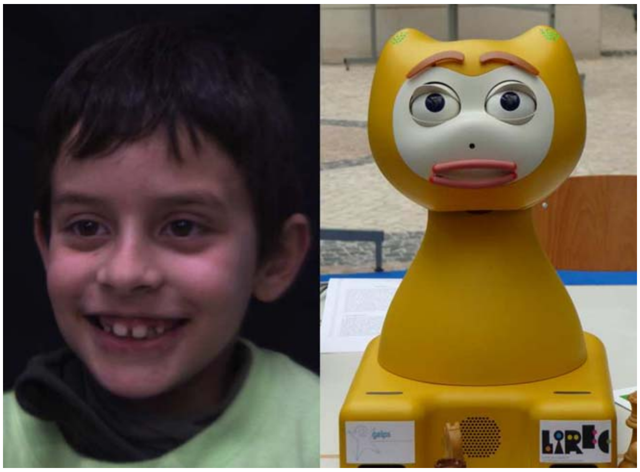

Fig. 1 Frontal view of a user playing chess with the iCat

After annotating the affective states, a similar process was adopted to annotate the non-verbal behaviours displayed by the children while interacting with the iCat. Based on their occurrence during the interaction with the robot, the following non-verbal behaviours were identified (see Table 1): *Looking at the iCat*, *Looking at the chessboard*, *Looking elsewhere*, *Smiling*, *Mouth fidget*, *Hand on mouth*, *Scratching face or head*, *Blinking while looking at the iCat*, *Raising eyebrows*, *Approaching*, *Moving away*. Contextual information such as the phase of the game was also considered to define when children were looking at the iCat. This process generated additional non-verbal behaviours for the annotation: *Looking at the iCat after own move*, when the iCat generates an affective reaction; *Looking at the iCat after iCat move*, when the user receives feedback from the

**Table 1** The annotated non-verbal behaviours

| Non-verbal behaviour |
|---|
| Looking at the iCat |
| Looking at the iCat after own move |
| Looking at the iCat after iCat move |
| Looking at the iCat during the game |
| Looking at the chessboard |
| Looking elsewhere |
| Smiling |
| Mouth fidget |
| Hand on mouth |
| Scratching face or head |
| Blinking while looking at the iCat |
| Raising eyebrows |
| Approaching |
| Moving away |

robot, such as approval or disapproval; and *Looking at the iCat during the game*, when the user is thinking and the iCat is performing idle behaviours such as blinking and looking sideways.

Two of the coders who annotated the user's states were also asked to annotate the portions of video segments in which the children exhibited these behaviours. Since the annotation of the non-verbal behaviours was performed after the annotation of the user's states, the latter was not affected by any previous observation of the user's expressions. As the coders were asked to annotate explicit behaviours displayed by the participants, annotations were almost identical (less than 1–2 frames difference). For each video segment each behaviour was assigned a value. This value was computed as the average number of frames over which a specific behaviour was displayed in the video segment.

### 3.3 Results and discussion

In order to investigate which non-verbal behaviours are affected by the states experienced by the user during the interaction with the iCat (i.e., *feeling* and *engagement with the iCat*) and how, statistical analysis was performed.

Two repeated measures $t$ tests ($N = 15$) were performed for each behaviour to explore whether there is a significant difference in the occurrence of the selected behaviour between *positive* and *negative* and between *engaged with the iCat* and *not engaged with the iCat* samples. In each test the type of behaviour was considered as the dependent variable, while the valence of the feeling or the engagement with the iCat as the independent variable (two levels).

**Positive versus negative**: Results showed a significant effect of the valence of the feeling on a subset of non-verbal behaviours: *Looking at the iCat*, *Looking at the iCat during the game*, *Looking at the chessboard*, and *Smiling*. Means and standard deviations are reported in Table 2.

In correpondence with positive feeling, overall children look at the iCat more than in correpondence with negative feeling [$t(14) = 3.84$; $p < 0.001$]. Also, a discrimination between positive and negative is possible when children look at the iCat during the game, which is more associated with positive feeling [$t(14) = 1.91$; $p < 0.05$]. Results also show that, in case of negative feeling, children tend to look more at the chessboard than in case of positive feeling [$t(14) = -5.03$; $p < 0.001$]. Finally, positive feeling is associated with a higher frequency of smiles compared with negative feeling [$t(14) = 4.51$; $p < 0.001$].

*Conclusion: the behaviours displayed by the children in the considered interaction scenario that are mainly affected by the feeling are the eye gaze and the smiles: when the feeling is positive, children tend to look at the iCat and smile more then when the feeling is negative.*

**Engaged with the iCat versus not engaged with the iCat**: Results showed a significant effect of the engagement with the iCat on the following behaviours: *Looking at the iCat*, *Looking at the iCat after own move*, *Looking at the chessboard*, *Smiling*, and *Blinking*. Means and standard deviations are reported in Table 3.

When children are engaged with the iCat, overall they tend to look more at the iCat than when they are not engaged with it [$t(14) = 5.88$; $p < 0.001$]. Taking into consideration the context in which the children look at the iCat, engagement and not engagement with the iCat appear particularly well discriminated based on the frequency of looking at the iCat after the user's move [$t(14) = 4.77$; $p < 0.001$]. Children appear to be looking at the chessboard more when they are not engaged with the iCat [$t(14) = -5.64$; $p < 0.001$] and to be smiling more when they are engaged with it [$t(14) = 2.68$; $p < 0.01$]. Blinking while looking at the iCat also appears to discriminate between the levels of engagement, with more blinking in the case of engagement with

**Table 2** Mean values and standard deviations of the non-verbal behaviours for the different conditions of feeling ($N = 15$ in each condition)

| Non-verbal behaviour | | Positive feeling | Negative feeling |
|---|---|---|---|
| Looking at the iCat | Mean | 0.69 | 0.21 |
| | S.D. | 0.28 | 0.33 |
| Looking at the iCat after own move | Mean | 0.39 | 0.18 |
| | S.D. | 0.42 | 0.33 |
| Looking at the iCat after iCat's move | Mean | 0.12 | 0.00 |
| | S.D. | 0.29 | 0.00 |
| Looking at the iCat during the game | Mean | 0.18 | 0.03 |
| | S.D. | 0.35 | 0.11 |
| Looking at the chessboard | Mean | 0.17 | 0.75 |
| | S.D. | 0.20 | 0.36 |
| Looking elsewhere | Mean | 0.14 | 0.04 |
| | S.D. | 0.23 | 0.12 |
| Smiling | Mean | 0.46 | 0.01 |
| | S.D. | 0.39 | 0.05 |
| Mouth fidget | Mean | 0.21 | 0.22 |
| | S.D. | 0.33 | 0.32 |
| Hand on mouth | Mean | 0.11 | 0.05 |
| | S.D. | 0.27 | 0.14 |
| Scratching face or head | Mean | 0.01 | 0.00 |
| | S.D. | 0.02 | 0.00 |
| Blinking while looking at the iCat | Mean | 0.02 | 0.00 |
| | S.D. | 0.04 | 0.01 |
| Raising eyebrows | Mean | 0.00 | 0.02 |
| | S.D. | 0.00 | 0.07 |
| Approaching | Mean | 0.03 | 0.00 |
| | S.D. | 0.12 | 0.00 |
| Moving away | Mean | 0.01 | 0.02 |
| | S.D. | 0.04 | 0.09 |

**Table 3** Mean values and standard deviations of the non-verbal behaviours for the different conditions of engagement with the iCat ($N = 15$ in each condition)

| Non-verbal behaviour | | Engaged with the iCat | Not engaged with the iCat |
|---|---|---|---|
| Looking at the iCat | Mean | 0.56 | 0.02 |
| | S.D. | 0.35 | 0.05 |
| Looking at the iCat after own move | Mean | 0.50 | 0.01 |
| | S.D. | 0.38 | 0.04 |
| Looking at the iCat after iCat's move | Mean | 0.01 | 0.00 |
| | S.D. | 0.02 | 0.00 |
| Looking at the iCat during the game | Mean | 0.06 | 0.01 |
| | S.D. | 0.22 | 0.03 |
| Looking at the chessboard | Mean | 0.39 | 0.96 |
| | S.D. | 0.36 | 0.09 |
| Looking elsewhere | Mean | 0.04 | 0.03 |
| | S.D. | 0.06 | 0.07 |
| Smiling | Mean | 0.25 | 0.00 |
| | S.D. | 0.35 | 0.00 |
| Mouth fidget | Mean | 0.24 | 0.23 |
| | S.D. | 0.33 | 0.34 |
| Hand on mouth | Mean | 0.07 | 0.04 |
| | S.D. | 0.19 | 0.12 |
| Scratching face or head | Mean | 0.04 | 0.00 |
| | S.D. | 0.15 | 0.00 |
| Blinking while looking at the iCat | Mean | 0.02 | 0.00 |
| | S.D. | 0.04 | 0.00 |
| Raising eyebrows | Mean | 0.01 | 0.02 |
| | S.D. | 0.03 | 0.07 |
| Approaching | Mean | 0.03 | 0.00 |
| | S.D. | 0.12 | 0.00 |
| Moving away | Mean | 0.01 | 0.00 |
| | S.D. | 0.04 | 0.00 |

the iCat [$t(14) = 2.20$; $p < 0.05$]: nevertheless, this result is not very significant, given that most of the times that the children look at the iCat they are engaged with it.

*Conclusion: eye gaze and smiles allows for a discrimination of the level of engagement with the iCat, with children looking at the iCat and smiling more when they are engaged with it.*

Overall, it seems that eye gaze and smiles play an important role in discriminating positive from negative feeling and engagement from not engagement with the iCat. Results also show that the information about the context in which the children look at the iCat can be useful for the purpose of discriminating among the identified states.

Other non-verbal behaviours did not prove significant in discriminating between *positive* and *negative* and between *engaged with the iCat* and *not engaged with the iCat*. This is expected, as many of them did not occur often during the interaction (e.g., hand on mouth, scratching face or head, raising eyebrows). Furthermore, others (e.g., mouth fidget) were continuously displayed throughout the interaction with the iCat, particularly by some of the users, and were not specific to any state. Finally, some of the behaviours, for example approaching and moving away, were difficult to capture from the annotators from the frontal view of the videos. Future work will investigate the role of the lateral postures of children using the view captured by the lateral camera.

## 4 Conclusion

This paper provides an overview of some of the issues and challenges in the design of an affect recognition framework for interactive artificial companions. As a requirement to allow for artificial companions to engage in a social interaction with human users, affect sensitivity is discussed with respect to key issues such as the ability to perceive spontaneous and subtle affective cues for the detection of affective states different from basic emotions, the ability to analyse multiple modalities of expression, the robustness to real world scenarios, the personalisation of the affect recognition abilities and the ability to adapt over time to changes of attitude of the user towards the companion.

Current findings in the domain of affect recognition represent a valuable resource for the design of affect sensitive companions. One of the drawbacks of this richness of studies in the field is that results and performances are often not comparable due to the different experimental conditions, the different databases of affective expressions used to train the affect detectors, the different data used, etc. Research on artificial companions should aim to establish common guidelines for the design of affect recognition frameworks suitable for real world applications.

In this paper we presented a real world scenario in which a robot companion plays chess with children. In this scenario, the non-verbal behaviours displayed by the children that allow for a better discrimination of application-dependent affective states related to the task and the social interaction with the robot are identified. Results show that the amount of time and the circumstances in which the user looks at the iCat, as well as the presence of smiles, can help discriminate between different types of feeling and levels of engagement with the iCat. When they experience a positive feeling, the children tend to smile more and look more at the iCat than when they experience a negative feeling. Looking at the iCat during the game, in particular, seem to discriminate well between the types of feeling. In case of engagement with the iCat, the children smile more and look more at the iCat (overall and when they look at the iCat after their own move) than when they are not engaged with it. The key role played by the contextual information combined to some of the non-verbal behaviours (i.e., looking at the iCat in this case) proved the need for a multimodal approach to affect recognition. These findings are in line with results from previous work, which showed that contextual information helps discriminate among different nuances of affective states in children playing with the iCat [43].

Finally, it is to be highlighted that this study is only exploratory, as it considered only a limited number of participants. Nevertheless, this work is not intended to be a comprehensive investigation of the relationship between the identified user states and the non-verbal behaviour displayed by the user. Rather, the main aim is to uncover insights to support the choices made in the first steps of the design of an affect recognition system. The reported results contribute to provide the foundations for the design of an affect recognition system for a game companion.

**Acknowledgements** This work was supported by the EU FP7 ICT-215554 project LIREC (LIving with Robots and intEractive Companions). The authors would like to thank Christopher Peters and Asimina Vasalou for providing useful comments.

## References

1. Breazeal C (2003) Emotion and sociable humanoid robots. Int J Hum-Comput Stud 59(1–2):119–155
2. Gratch J, Wang N, Gerten J, Fast E, Duffy R (2007) Creating rapport with virtual agents. In: 7th international conference on intelligent virtual agents, Paris, France
3. Dautenhahn K (2007) Socially intelligent robots: dimensions of human-robot interaction. Philos Trans R Soc B Biol Sci 362(1480):679–704
4. Zeng Z, Pantic M, Roisman GI, Huang TS (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 31(1):39–58
5. Kapoor A, Burleson W, Picard RW (2007) Automatic prediction of frustration. Int J Hum-Comput Stud 65(8):724–736
6. El Kaliouby R, Robinson P (2005) Generalization of a vision-based computational model of mind-reading. In: 1st international conference on affective computing and intelligent interaction, Beijing, China
7. Yeasin M, Bullot B, Sharma R (2006) Recognition of facial expressions and measurement of levels of interest from video. IEEE Trans Multimedia 8(3):500–507
8. Peters C, Asteriadis S, Karpouzis K, de Sevin E (2008) Towards a real-time gaze-based shared attention for a virtual agent. In: Workshop on affective interaction in natural environments (AFFINE), ACM international conference on multimodal interfaces (ICMI'08), Chania, Crete, Greece
9. Castellano G, Pereira A, Leite I, Paiva A, McOwan PW (2009) Detecting user engagement with a robot companion using task and social interaction-based features. In: International conference on multimodal interfaces and workshop on machine learning for multimodal interaction (ICMI-MLMI'09). ACM Press, Cambridge
10. Caridakis G, Malatesta L, Kessous L, Amir N, Raouzaiou A, Karpouzis K (2006) Modeling naturalistic affective states via facial and vocal expression recognition. In: International conference on multimodal interfaces, pp 146–154
11. Castellano G, Mortillaro M, Camurri A, Volpe G, Scherer K (2008) Automated analysis of body movement in emotionally expressive piano performances. Music Percept 26(2):103–119
12. Littlewort GC, Bartlett MS, Lee K (2007) Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain. In: International conference of multimodal interfaces, pp 15–21
13. Scherer KR (2000) Psychological models of emotion. In: Borod J (ed) The neuropsychology of emotion. Oxford University Press, Oxford, pp 137–162
14. Bänziger T, Scherer K (2007) Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus. In: 2nd international conference on affective computing and intelligent interaction, Lisbon
15. Ioannou S, Raouzaiou A, Tzouvaras V, Mailis T, Karpouzis K, Kollias S (2005) Emotion recognition through facial expression analysis based on a neurofuzzy method. Neural Netw 18:423–435
16. Devillers L, Vasilescu I (2006) Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In: International conference on spoken language processing
17. Valstar MF, Gunes H, Pantic M (2007) How to distinguish posed from spontaneous smiles using geometric features. In: ACM international conference on multimodal interfaces (ICMI'07), Nagoya, Japan, pp 38–45
18. Gunes H, Piccardi M (2009) Automatic temporal segment detection and affect recognition from face and body display. IEEE Trans Syst Man Cybern, Part B 39(1):64–84
19. Kim J, Andre E, Rehm M, Vogt T, Wagner J (2005) Integrating information from speech and physiological signals to achieve emotional sensitivity. In: 9th European conference on speech communication and technology
20. Castellano G, Kessous L, Caridakis G (2008) Emotion recognition through multiple modalities: face, body gesture, speech. In: Peter C, Beale R (eds) Affect and Emotion in Human-Computer Interaction. LNCS, vol 4868. Springer, Heidelberg
21. Meeren H, Heijnsbergen C, Gelder B (2005) Rapid perceptual integration of facial expression and emotional body language. Proc Natl Acad Sci USA 102(45):16518–16523
22. Stein B, Meredith MA (1993) The merging of senses. MIT Press, Cambridge
23. Shan C, Gong S, McOwan PW (2007) Beyond facial expressions: learning human emotion from body gestures. In: Proceedings of British machine vision conference (BMVC'07), Warwick, UK
24. Zeng Z, Hu Y, Liu M, Fu Y, Huang TS (2006) Training combination strategy of multi-stream fused hidden Markov model for audio-visual affect recognition. In: ACM international conference on multimedia, pp 65–68
25. Pantic M, Patras I (2006) Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE Trans Syst Man Cybern, Part B 36(2):433–449
26. Anderson K, McOwan PW (2006) A real-time automated system for recognition of human facial expressions. IEEE Trans Syst Man Cybern, Part B 36(1):96–105
27. Castellano G, Villalba SD, Camurri A (2007) Recognising human emotions from body movement and gesture dynamics. In: 2nd international conference on affective computing and intelligent interaction, Lisbon
28. Scherer KR (1984) On the nature and function of emotion: a component process approach. In: Scherer KR, Ekman P (eds) Approaches to emotion. Erlbaum, Hillsdale, pp 293–317
29. Kapoor A, Picard RW (2005) Multimodal affect recognition in learning environments. In: ACM international conference on multimedia, pp 677–682
30. You Z-J, Shen C-Y, Chang C-W, Liu B-J, Chen G-D (2006) A robot as a teaching assistant in an English class. In: Sixth international conference on advanced learning technologies

31. Dias J, Paiva A (2005) Feeling and reasoning: a computational model for emotional characters. In: Bento C, Cardoso A, Dias G (eds) Progress in artificial intelligence, EPIA'2005. LNAI, vol 3808. Springer, Berlin
32. Castellano G (2008) Movement expressivity analysis in affective computers: from recognition to expression of emotion, PhD thesis, Department of Communication, Computer and System Sciences, University of Genova, Italy
33. Graham S, Weiner B (1996) Theories and principles of motivation. In: Berliner DC, Calfee RC (eds) Handbook of educational psychology. Macmillan, New York, pp 63–84
34. Arroyo I, Cooper DG, Burleson W, Woolf BP, Muldner K, Christopherson R (2009) Emotions sensors go to school. In: International conference on artificial intelligence in education, Brighton, UK, pp 17–24
35. Horgan DD, Morgan D (1990) Chess expertise in children. Appl Cogn Psychol 4(2):109–128
36. Waters AJ, Gobet F, Leyden G (2002) Visuospatial abilities of chess players. Br J Psychol 93(4):557–565
37. Linder I (1990) Chess, a subject taught at school. Sputnik: Digest of the Soviet Press, pp 164–166
38. Breemen A, Yan X, Meerbeek B (2005) iCat: An animated user-interface robot with personality. In: Pechoucek, Steiner, Thompson (eds) Proceedings of autonomous agents and multiagent systems conference, AAMAS'05. ACM Press, New York, pp 143–144
39. Leite I, Pereira A, Martinho C, Paiva A (2008) Are emotional robots more fun to play with? In: Proceedings of IEEE RO-MAN 2008 conference, Munich, Germany
40. Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39:1161–1178
41. Poggi I (2007) Mind, hands, face and body. A goal and belief view of multimodal communication. Weidler, Berlin
42. Kipp M (2008) Spatiotemporal coding in ANVIL. In: Proceedings of the 6th international conference on language resources and evaluation (LREC-08)
43. Castellano G, Leite I, Pereira A, Martinho C, Paiva A, McOwan PW (2009) It's all in the game: towards an affect sensitive and context aware game companion. In: International conference on affective computing and intelligent interaction, Amsterdam, The Netherlands. IEEE Press, New York