Emotion Modelling for Social Robots

Ana Paiva, Iolanda Leite and Tiago Ribeiro
INESC-ID and Instituto Superior Técnico, Technical University of Lisbon
Av. Prof. Cavaco Silva, 2744-016, Oeiras, Portugal
{ana.paiva, iolanda.leite, tiago.ribeiro}@inesc-id.pt

Abstract

This chapter describes current advances in emotion modelling for social robots. It begins by contextualising the role of emotions in social robots, considering the concept of the Affective Loop. It describes a number of elements for the synthesis and expression of emotions through robotic embodiments and provides an overview of the area of emotional adaptation and empathy in social robots.

Keywords: Emotion Modelling, Social Robots, Human-Robot Interaction, Affective Loop.

1. Introduction: Robots in the Affective Loop

The concept of a self-operating machine that resembles humans and behaves similarly to humans dates to ancient civilizations. From Homer, Leonardo da Vinci and Isaac Asimov, robots have captured the imagination of philosophers and writers throughout history, and robots continue to inspire us. Robotics is a technology that is expected to change the world and the way we live. During the early stages of robotics research, much of the development was focused on the usefulness of robots in industrial settings. As robots progress from these very controlled settings and laboratory environments and are deployed in homes and social contexts, the ability of robots to interact with humans in ways that resemble human interaction becomes increasingly more relevant (Breazeal, 2009). Emotions are essential for that interaction. To portray emotions in robots, researchers require methods to model and express emotions with different embodiments and distinct manners. Such modelling of emotions allows the placement of robots in the context of the Affective Loop to foster social interaction. As defined by Höök (2009), the *Affective Loop* is the interactive process in which “the user [of the system] first expresses her emotions through some physical interaction involving her body, for example, through gestures or manipulations; and the system then responds by generating affective expression, using for example, colours, animations, and haptics” which “in turn affects the user (mind and body) making the user respond and step-by-step feel more and more involved with the system.”

To establish this Affective Loop between users and robots (Figure 1.1), robots will require an affect detection system that recognises, among other states, whether the user is experiencing positive or negative feelings, and a reasoning and action selection mechanism that chooses the optimum emotional response to display at a cognitive level. The method by which robots express the intended affective states should be effective (to be perceived by the users), and the actions of the emotional robot will affect the user (the third step of the Affective Loop). The robot perceives the user with the goal of personalising the interaction by analysing the user’s responses to the various affective expressions of the robot and adapting its emotional behaviour for each particular user.

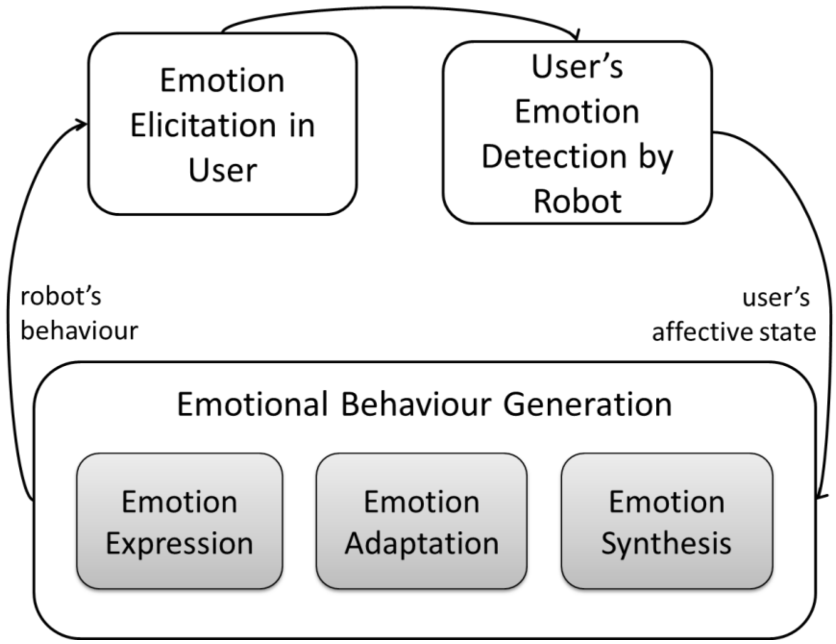

Figure 1.1. Affective Loop of Emotional Robots.

Affective interactions play different roles and have various purposes in the context of Human-Robot Interaction (HRI). Among others, we can distinguish the following:

1 *Give the illusion of life* - The design of adaptive emotional behaviour must use particular caution to avoid unexpected or unintelligible behaviour. This problem can be solved by following a number of guidelines on the methods for creating expressive behaviour in robots, which provide the robots with the "illusion of life" (Ribeiro & Paiva, 2012). This illusion will lead to the user's "suspension of disbelief," which increases the perception of social presence, thus rendering the robot as a believable character (Bates, 1994).

2 *Augment engagement* - Emotions contribute to engagement in a social interaction context. Engagement, in this context, is defined as "the process by which two (or more) participants establish, maintain and end their perceived connection" (Sidner et al., 2004) and has received increasingly more attention by the HRI community (Rich et al., 2010). As previous research has highlighted, appropriate displays of affect have a significant effect on the user's engagement while interacting with social robots (Leite et al., 2012).

3 *Augment social presence in the long-term* - The lack of adaptive emotional behaviour decreases the user's perception of social presence, especially during long-term interactions (Leite et al., 2009), which in turn renders the robots to be non-believable characters (Bates, 1994). To be perceived as socially present, social robots must not only convey believable affective expressions, but also be able to do so in an intelligent and personalised manner, for example, by gradually adapting their affective behaviour to the particular needs and/or preferences of the users.

This chapter will discuss the modelling of emotions in robots, not only through explicit computational mechanisms in which the emotions are captured but also in terms of adaptation, personalisation and expression, leading our emotional robots to become empathic, emotional creatures that can sustain the Affective Loop with the user.

2. Creating Synthetic Emotions in Robots

Robots possess the power to convey the illusion of life simply by their physical presence and simple movements. When a robot moves towards a door and suddenly backs up, one may interpret its actions as avoidance and fear. Our perception of the robots' actions may be biased, or perhaps enriched, by our inclination to suspend our disbelief and see robots as intelligent creatures that act according to their desires, goals and emotions. However, a robot's behaviour can result from computational mechanisms that do not explicitly capture any of those aspects. Significant work on emotional behaviour in robots has been driven by the fact that simple behaviour may "fool us" by leading us to perceive robots as having "emotions" when, in reality, those "emotions" are not explicitly modelled and simply arise as an emergent effect of specific simple patterns of behaviour.

When situations are such that a robot may need to interact with users at a higher level, often using natural language and gestures, and their actions need to be rich enough to convey some goal-oriented behaviour, emotions may represent a way to model different responses and thus provide the robot with more believable and appropriate responses to the tasks. In those situations, emotions may represent abstract constructs that simplify the generation of robots' behaviours, as described by Leite et al. (2013). In this case, emotion modelling in robots becomes explicit, and specific architectures for that modelling have emerged in recent years. Emotions, modelled explicitly, may affect not only the action selection but also other cognitive processes such as reasoning, planning and learning. As this area grows, new types of architectures exploring different mental processes (e.g., theory of mind or affect regulation) are also arising, allowing robots to perform better in the world and interact with humans in a more natural way.

In general, emotional architectures for robots (and virtual agents) seek inspiration from the way humans and other species perceive, reason, learn and act upon the world. We may distinguish different types of architectures according to their inspiration (from a neurobiological inspiration, to more psychological or data-driven models); the affective states they try to model (e.g., emotions, moods, personality); the types of processes captured (e.g., appraisal, coping); the integration with other cognitive capabilities; and the expressive power they possess. Most of the existing types of architectures are built with different processes and levels, thus extending a generic hybrid model. As argued by A. Sloman (2002), different types of architectures will support distinct collections of states and processes. A number of these types rely on symbolic models to represent the perceptual elements of the world, whereas others take a non-symbolic approach based on neural modelling.

Considering the neurobiological approaches, they generally take the view that modelling of emotions can be done through different computational constructs associated with structures from the central nervous system such as the amygdala and the hypothalamus (Arbib, 2004). One of the earlier types of architectures for emotions in robots was Cathexis (Velazquez, 1998), which is a computational model of emotions and action selection inspired by neurobiological theories. The architecture integrates drives and emotions in a way that guides the behaviours and decision making of the robot.

Many models capture affective states in an emergent manner, as the resulting pattern of behaviour arising from a variety of different processes embedded in the agent. More explicit modelling is accomplished by representing affective states using a symbolic approach. The classic BDI reference model (Wooldridge, 1995) considers "Beliefs, Desires and Intentions" to be the basic mental attitudes for generating an agent's intelligent behaviour. This reference model has been extended to capture other mental states, in particular, emotions (see, for example, Gratch & Marsella, 2004 and Dias & Paiva, 2005). The majority of these types of architectures focus primarily on representing emotional states (e.g., quickly active, short and focused states such as anger or joy), in particular the six basic emotional states (e.g., anger, joy, surprise, sadness, disgust and fear), as hypothesised by Ekman & Friesen (1975). Other affective states have been considered, such as moods (for example by Leite et al., 2012a and Álvarez 2010), dispositions, sentiments, or personality (Tapus et al., 2008; Woods et al., 2005).

The construction of the models mentioned above rely on theories of specific emotion, which are adopted as the basis for a computational system and thus dictate how an emotional state is triggered in a robot. One of the most influential theories, proposed by Magda Arnold in the 1960s, considers that emotions arise from the appraisal of situations and that an emotion is thus a result of the evaluation of the interconnection between the self and the environment and its objects (see Chapter 5 in this volume for more details on appraisal models of emotion). Appraisals result from an attraction towards or repulsion from objects, which in turn give a valence to the emotion. In robots, we can imagine the process of emotion generation as resulting from the subjective evaluation of the situation that the robot is facing. If a robot is presented with a situation that is good for itself (for example, if it is playing a game and is at an advantage), the situation should lead to an internal process that will create specific variables to attain values that lead to the triggering of the emotional status of joy.

The process of appraising the situation has been treated differently in various types of emotion architectures for robots. In the iCat chess player (Leite et al., 2008), a robot that provides affective feedback to the user, emotions result from affective signals that emerge from an anticipatory system containing a predictive model of itself and/or of its environment (Martinho & Paiva 2006). This anticipatory system generates an affective signal resulting from the mismatch between what is expected and what the robot senses. If the robot expects the user to perform well in the game and the user makes a mistake, it is an unexpected and positive (for the robot) situation leading to the generation of a positive valence affective signal.

Emotions condition and structure our perceptions, direct our attention and prepare us for action. The signalling in the nervous system that occurs from the primary appraisal processes is key to emotional priming and thus to our perception of emotion. Robots are situated in and rely strongly on their perceptions from the environment to act upon that environment. In an emotional robot, not only are its perceptual processes affected by its emotional states, but a number of emotional states may also arise directly from those "untreated" perceptions. For example, Kismet (Breazeal, 2003) possessed a quite advanced perceptual system and evaluated external stimuli according to the robot's context in terms of its affective state and drives.

Emotions affect behaviour. They are associated with different types of actions, such as fight or flight responses. In his seminal work, N. Frijda (1986) argued that our autonomic activity, the core of our emotional states, leads to states of “action readiness”. As a response to the environment, emotions prepare the individual for action. Many different types of action tendencies have been studied and can be captured in our agent architectures. As such, one should consider the ways in which the affective states influence the actions of the agents and which types of internal mechanisms are considered in the process of action selection, considering emotional states. Behaviour generation was done in the iGrace system, which uses a database of emotional experiences as the basis for action.

3. Emotional Expression in Robots: the Illusion of Robotic Life

It is necessary for a social robot to model and select actions based on affective goals and conditions, and a robot must be able to properly express those actions with an adequate affective display. This expression of emotion can be observed in the Affective Loop as the connection point between Emotional Behaviour Generation (resulting from the use of an emotional model) and Emotion Elicitation in the user. How should one generate expressive behaviour in robots that could be interpreted correctly by humans? How can such expressive behaviour be generated in a way that is flexible and, as much as possible, transposed between different embodiments?

In terms of concrete emotions, studies from psychology can inform us about how humans interpret emotions and their expression. Authors such as Ekman and Friesen (1975), Russell (1994) and Elfenbein and Ambady (2002) have studied and published papers about how the human ability to recognize emotions in other humans is universal. If it were possible to develop robotic characters able to process and express emotions as humans do, then these links to psychology would most likely be the best starting point. The majority of current robots are very limited in terms of expression, and many of them lack a number of expressive features that humans have, such as a mouth or facial muscles. This has led many researchers to turn to the arts in search of inspiration and solutions.

3.1 Principles of animation for expressing emotions in robots

The influential work by Bates (1994) was one of the first to seek inspiration from the arts, in particular from *The Illusion of Life*, a well-known book by Frank Thomas and Ollie Johnston, in which they describe over 60 years of experience in creating characters at the Walt Disney Animation Studios (Thomas & Johnston, 1981). The book intrinsically refers to the Affective Loop - the way in which Disney characters are brought to (an illusion of) life, by enacting stories which “make people laugh - and even cry.” Other authors have followed the identical philosophy, including Reilly (1996) who demonstrated that “it is artists who best know how to create believable characters and how to imbue them with emotions.” More recently, van Breemen (2004) has taken this concept into the expressive robotics field by claiming that “user-interface robots have the same problem as the early day of animation: they miss the illusion of life.”

These ideas lead the HRI community to look at one key concept for social robotic characters - the concept of the robot as a “believable character”. Bates (1994) defines a believable character as “not an honest or reliable character, but one that provides the illusion of life, and thus permits the audience’s suspension of disbelief.” The core of Thomas & Johnston’s *The Illusion of Life* is composed of Disney's Twelve Principles of Animation. These principles are presented in the book and serve as rules for professional animators as guidelines on how to create believable and expressive characters. The reader can refer to the book to understand each of the twelve principles from the authors, who were Disney animators.

The attempt to apply these principles of animation to robots has continued for almost a decade. Several authors have proposed this method, starting with van Breemen (2004), who defined robot animation as “the process of computing how the robot should act such that it is believable and interactive”. His attempt was to look at a number of the principles to understand how to create better expressions for the iCat robot, and he implemented a method that he named Merging Logic, which corresponds to the principle of Slow In/Slow Out.

Wistort (2010) has looked at the Disney principles to discuss how a number of them could be applied to robots. He emphasises the need for robots to have expressive eyes and insists that actuated pupils help convey life through the eyes. He stresses the need to have silent eyes because most robots rely on motors to move their eyes, and these are usually quite noisy. This work is interesting because it tries to bend the definition of the principles with a more scientific understanding of how the principles actually work. That is a necessary step because the traditional principles help explain the methods that animators should use to design animation. When considering social robots, we need to understand how such principles can be adapted or extended to work under autonomous animation.

Takayama et al. (2011) used Anticipation, Engagement, Confidence and Timing to enhance the readability of a robot’s actions. They refer to the definition of “thinking characters” by Lasseter (1987), in which “the animator gives life to the character by connecting its actions with a thought process.” One especially interesting distinction they demonstrate is about Functional Motions versus Expressive Motions. As robots are part of our physical world, it is likely that they will perform a blend of these two types of motions. A simple locomotive behaviour may carry both types of motion: Functional Motion to enable the robot to move from one place to another and Expressive Motion to express how the robot is feeling. Speaking to or requesting something from someone is functional and expressive. In a number of situations, we will want to separate them, as when a robot first looks at a door handle (expressive motion) and then grasps it to open the door (functional motion). By looking at the door handle, the robot first expresses interest and attention towards the handle, so humans can anticipate that it is going to grasp the handle and open the door.

Mead and Mataric (2010) used a number of principles of animation to improve the understanding by autistic children of a robot’s intentions. For “Exaggeration”, they were inspired by a process of creating caricatures by exaggerating the difference from the mean. The features that uniquely identify an expression are isolated, and those features are amplified to make them even more relevant, thus producing exaggerated expressions. Gielniak et al. (2012) followed a similar idea but addressed it as a motion signal processing technique. They successfully developed an algorithm that creates exaggerated variants of a motion in real time by contrasting the motion signal, and they demonstrated this idea by applying it to their SIMON robot.

In our recent work, we examined this issue from a holistic point of view by providing insight on how each of the traditional principles of animation could be used in robotics and how they could be used to enhance the readability of Ekman's six basic emotions in the EMYS robot (Ribeiro & Paiva, 2012). The majority of the traditional principles of animation seem to apply to robot animation.

There is much to understand about how the traditional principles of animation can be applied to robots. The major difference that we find is that they are not only robots, but they are also interactive, and Disney's traditional principles were not intended for interactive characters. This difference does not mean that we cannot apply the traditional principles, but it does convert the issue into an engineering problem. Slow-In/Slow-Out, Arcs and Exaggeration may be easily incorporated into an animation system by warping the motion signals. However, what about the other principles? How do we programmatically define the use of correct Timing?

3.2. Computing Emotional Expressions in Robots

To support the generation of complex expressive behaviour in robots, we need a flexible and consistent architecture. The robotics community, influenced by the work on conversational characters, adopts generic frameworks such as the SAIBA (Situation, Agent, Intention, Behaviour, Animation) framework (Kopp et al., 2006). This framework, illustrated in Figure 3.1, is divided into three phases:

Intent Planning produces a Functional Markup Language (FML), which contains a definition of that which the character has intended to do.

Behaviour Planning receives the FML and produces a Behaviour Markup Language (BML) (Vilhjalmsson et al., 2009), which contains the specific details about the manner in which the character is planning to perform the behaviour generally specified in the FML.

Behaviour Realisation receives the BML, interprets it regarding the actual character that will perform it, and applies the actual behaviours onto the character. In an analogy to film, it can be viewed as the actor performing and following the orders of the Director and Casting Director.

Figure 3.1: The SAIBA framework.

Although SAIBA was initially intended for virtual characters, a number of authors have applied it to robot animation because BML is not character-specific (meaning that an identical BML script can be used by different Realisers). It has been used to make the GRETA animated agent interact with an AIBO robot (Moubayed et al., 2008) and a NAO robot (Niewiadomski et al., 2011). We have developed a system based on BML that allows continuous interaction with robots by providing mechanisms that can cause external events (perceptions) to interact with pre-defined BML sequences. Following the trend of Niewiadomski et al. (2011) and Kipp et al. (2010), we divided the Behaviour Realisation phase into the Behaviour Scheduling and Body Execution sub phases. The system was demonstrated with an interaction between an EMYS robot and a NAO robot, with external audio, visual and tangible events (Ribeiro & Paiva, 2012).

A popular middleware that was developed by Willow Garage[1] is ROS - Robot Operating System (Quigley et al., 2009). Although it actually works as a communication layer, there has been a growing development of the robots and modules that run it. By having a ROS-compliant robot, one can use any of the ROS modules that have been made freely available and can be used especially for navigation, vision, or arm manipulation. Following Holroyd & Rich (2012), who recently developed a BML module for ROS, other authors may develop expressive modules for ROS that could be shared throughout the robotics community.

3.3. Beyond Traditional Emotion Expression in Robots: Using Form, Colour and Sound

If we are able to develop a model that can compute an emotion and the intention of its display, we must map that to the expressive features of a robot. With as many diverse embodiments as are illustrated in Figure 3.2, it is very difficult to generalise on how this mapping could be done.

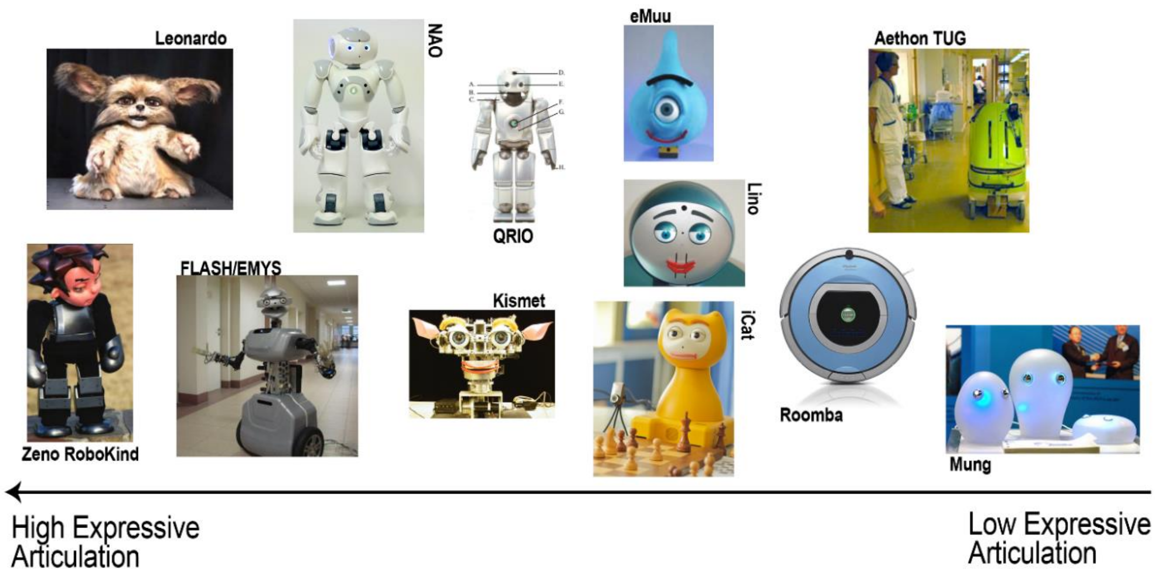

Figure 3.2: Multitude of robotic embodiments along a dimension of Expressive Articulation. Robots on the left contain more degrees of freedom available for expressivity.

[1] http://www.willowgarage.com/

On the left side of Figure 3.2, with Mid to High Expressive Articulation, we first highlight the work using eMuu (Bartneck, 2002), Lino (Krose et al., 2003), EMYS (Ribeiro & Paiva, 2012), Kismet (Breazeal, 2003), the iCat (Breemen, 2004; Leite et al., 2008) and Leonardo (Breazeal et al., 2004). These robots are not humanoids, but they perform emotional expression through the use of many degrees of freedom to try to mimic human expressions (in a simplified and iconic way).

With respect to humanoids, there have been studies on emotional expression, especially with the QRIO robot (Tanaka et al., 2004) and the NAO robot (Beck et al., 2010). The majority of humanoid robots have very simple faces and, as such, rely solely on body expression - which is more difficult to model and more ambiguous to read. Hanson Robotics has recently developed the Zeno RoboKind (Hanson et al., 2009), which features a highly articulated body and a highly expressive face based on muscular features. This robot introduces an accessible and flexible platform that contains facial and bodily expression, but there are no studies on its expressivity yet.

Focusing on the right side of Figure 3.2 presents an issue. Although social robots are generally designed to be expressive, there are some types of robots (social or not) that do not display expressive features. As Bethel (2009) notes, search-and-rescue robots are designed to move across disaster zones to find and reach victims. Some socially assistive robots may be designed to help elderly or disabled people get out of bed and move around, and as such, their purpose is to be steady and move safely, as does the RobCab robot (Ljungblad et al., 2012). In a domestic environment, we may have functional robots such as Roomba[2] (robotic vacuum cleaner) or smaller robots with a minimal design that are intended to behave as personal companions and assistants, such as the Mung robot (Kim et al., 2009). These more simplistic robots may rely on sound and light to augment their expressivity. Using sound and light can actually be considered part of the traditional "Staging" principle of animation because it refers to focusing on the elements that are important. One of the ways to achieve that is by properly using sound, music, lights and shadows. By re-defining this principle into "Intention", we do not invalidate these practices because it helps focus on the clear meaning of that which the robot is expressing. Recently, a number of researchers have augmented the expressivity of a Roomba robot by adding expressive lights (Rea et al., 2012) and an expressive tail (Singh & Young, 2012) (Figure 3.3). Other researchers have added a projector to a telemedicine robot that could project signals on the ground to demonstrate the robot's intention (Shindev et al., 2012).

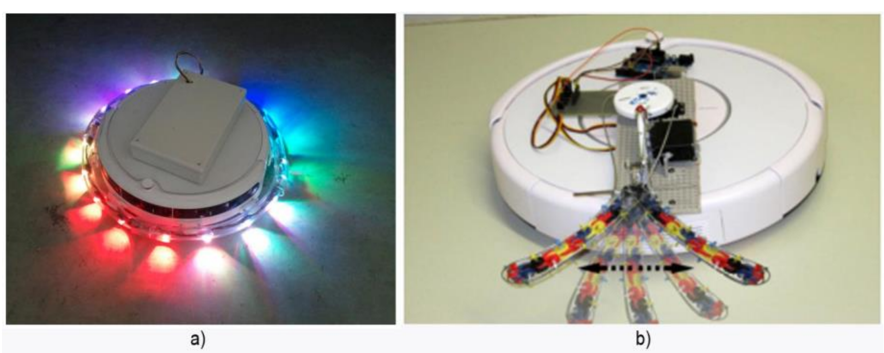

Figure 3.3: (a) A Roomba robot modified to include expressive lights (Rea et al., 2012); (b) A Roomba robot modified to include an expressive tail (Singh & Young, 2012).

[2] http://www.irobot.com

4. Emotion Adaptation in Robots: Towards Empathic Robots

An emotional robot should appraise its surrounding environment and react emotionally to it, and it should perceive other robots' and the user's affective states. Such perception and sensitivity of the other's affective states is essential to place the robot in the Affective Loop. A robot should perceive the user's affective state and act in response to such perceptions. But how can the robot close the loop and empathise with the user by providing more appropriate and adaptive responses during the interaction?

Empathy can be defined as "an affective response more appropriate to someone else's situation than to one's own," which can result from mechanisms such as motor mimicry, classical conditioning, direct association of cues from the person with whom we are empathising, mediated association of cues through semantic processing, and perspective-taking (Hoffman, 2001). Hoffman considers the first three mechanisms to be automatic and involuntary, whereas the last two are higher-order cognitive modes. The process of empathy can go beyond a merely affective response in tune with the affective state of the person with whom we are empathising. Empathy has been considered one of the major determinants of prosocial actions (Feshbach, 1978; Eisenberg & Fabes, 1990; Davis, 1994), i.e., "voluntary behaviour intended to benefit another such as helping, donating, sharing and comforting" (Eisenberg et al., 1997). In the literature, prosocial behaviours are often designated as socially supportive behaviours (Cutrona et al., 1990). These behaviours are more likely to be triggered when observing someone in pain, danger or other types of distress (Hoffman, 2001).

In the virtual agents community, empathy is a research topic that has received much attention in the last decade (Paiva et al., 2004; Prendinger & Ishizuka 2005; Bickmore & Picard, 2005; Brave et al., 2005; McQuiggan & Lester, 2007). In HRI, the research addressing empathy has recently appeared. One possible reason for this delay is that robots share our physical space, and recognising the user's affective state is thus a more challenging task because the interactions tend to be more open-ended. When users interact with virtual agents, they are usually in front of a computer screen, and there is frequently the possibility of selecting predefined dialogues or actions to inform the agent about the user's affective state (Bickmore & Picard, 2005). Recent advances in automatic affect recognition using different modalities, including vision, speech or physiological signals, contribute significantly to overcoming this obstacle (for a review, see Zeng et al., 2009). Many important questions remain unanswered: Will social robots benefit for being empathic? Will interactions become more natural and enjoyable if robots can adapt their behaviour to the user's affective state? What is the state of the art on empathic robots?

A significant segment of the research addressing empathy in social robots has focused on automatic and involuntary processes (as defined by Hoffman), such as mimicking the user's affective state. Hegel et al. (2006) presented a study with an anthropomorphic robot that recognises the user's emotional state through speech intonation and mirrors the inferred state using a corresponding facial expression. The results suggested that the users who interacted with this version of the robot found the robot's responses to be more adequate in their appropriateness to the social situation and timing than the users who interacted with the robot without affective expressions. In another study (Riek et al., 2010), a robot in the form of a chimpanzee's head mimicked the user's mouth and head movements. After interacting with this robot, most of the subjects considered the interaction to be more satisfactory than participants who interacted with a version of the robot without the mimicking capabilities.

Perspective taking is considered one of the most “empathy-arousing” mechanisms in humans. The process of taking the perspective of others and attributing mental states such as thoughts, desires, knowledge and intentions that are not directly observable is known as Theory of Mind (Premack, 1978; Baron-Cohen, 1997). This capability has been considered an essential component of empathy by several authors (Feshbach, 1978; Baron-Cohen et al., 2009), especially because of its importance in emotion recognition (Adolphs 2002; Heberlein et al., 2009). Appraising the context of a task can be computationally more reliable than the affect recognition mechanisms based on other modalities used today, and perspective taking has also been used in HRI. Cramer et al. (2010) used perspective taking to drive the behaviour of a social robot in a study that was conducted to assess how empathy affects human attitudes towards robots. Using a between-subjects design, two groups of participants saw a four-minute video with an actor playing a cooperative game with an iCat robot. In one condition, the robot accurately expressed empathic behaviour toward the actor, whereas in the other condition, the robot inaccurately expressed empathic behaviour toward the actor (i.e., incongruent to the situation). Because the robot varied its empathic behaviour according to the performance of the other player in the game, we say that it was taking the perspective of the human in the scenario. In this study, there was a significant negative effect on the participants’ trust in the inaccurate empathic behaviour condition. Conversely, the participants who observed the robot displaying accurate empathic behaviours perceived their relationship with the robot as closer than those participants who observed the robot inaccurately expressing empathic behaviour.

In our previous work (Leite et al., 2013), we presented an empathic model based on perspective taking. To evaluate this model, we developed a scenario in which an iCat robot acts as a social companion to two players in a chess game. The robot reacted to the moves played on the chessboard by displaying several facial expressions and verbal utterances, showing empathic behaviour towards one player and behaving neutrally towards the other. The empathic responses were modulated by the relationship with the players (companion or opponent). The results of this study indicate that the users towards whom the robot behaved empathically perceived the robot as friendlier, suggesting that empathy plays a key role in HRI. There are a number of examples of perspective taking in HRI with the goal of improving a robot’s effectiveness in learning simple tasks from ambiguous demonstrations (Breazeal et al, 2006) or facilitating human-robot communication (Torrey et al., 2010).

In cases in which a social robot is able to display empathy by changing its behaviour according to the affective state that the user is experiencing, one can consider that adaptation to the user is occurring. There are several types of empathic responses that a person (or, eventually, a social robot) can take, from affective reactions to socially supportive behaviours. A relevant question arises as to how should robots select appropriate empathic responses for a particular user.

The answer to this question poses another opportunity for personalisation. Because the theoretical literature on empathy does not clearly specify which types of empathic responses are more likely to be triggered by certain affective states, the robot can adjust its empathic behaviour based on the preferences of particular users. In a recent empirical study, we found that the presence of adaptive empathic behaviours in a social robot resulted in the perception of the robot by children as being socially supportive in a similar extent to the way in which they feel about their peers (Leite et al., 2012a). In this study, the robot was able to use a no-regret Reinforcement Learning (RL) algorithm to learn the types of supportive behaviours that were more effective for each child, i.e., the behaviours that were more likely to increase a positive affective state in the child. The reward function of the RL algorithm was based on the non-verbal responses of the child captured a few seconds after the robot employed each supportive strategy. At each time step, no-regret algorithms make the best use of the feedback received so far. A policy that allows the iCat to select the next supportive behaviour by considering the previous history of employed behaviours and consequent rewards was implemented to learn the most effective supportive behaviours for a particular user and adapt to the interaction accordingly. The intuition that is the basis of this policy is simple: the robot selects the supportive behaviours with higher average rewards more often because they have higher probabilities. There is a possibility that other behaviours could be selected because although they had lower rewards in the past, the user may find them more suitable in the future.

Other researchers have explored the user's affective state as a mechanism to adapt the robot's behaviours. Liu et al. (2008) developed a scenario in which the robot learns the most appropriate behaviours by accounting for the physiological signals of the children. A preliminary study conducted with six autistic children showed that the robot was able to select appropriate behaviours based on the attraction levels expressed by the children collected through physiological signals. The question of adaptation has received particular attention in robot-assisted therapy, particularly in interaction with children with autism spectrum disorders because they require unique capabilities that non-adaptive robots might be unable to fulfil (Dautenhahn et al., 2003). Bekele et al. (2011) developed a head tracker that allows a NAO robot to adapt its postures, thus encouraging cooperative attention between autistic children and the robot.

Gonsior et al. (2012) conducted a study in which a social robot adapts its own mood to the user's mood in an attempt to increase empathy and helpfulness. The robot adjusts its nonverbal and verbal behaviours considering the user's self-assessments of affect. The results show that the subjects who interacted with the adaptive version showed a significantly higher desire to help the robot in further tasks than the participants in the control group, who interacted with a version of the robot with emotional behaviour but without adaptation capabilities.

Considering the work presented in this section, we hypothesise that empathy and, in particular, the consequent user adaptation that empathy facilitates will most likely make social robots more flexible and personalised to particular users, in the same way as other types of adaptation have been reported to be facilitators of Human-Robot Interaction (Koay et al., 2007; Lee et al., 2012). As stated by Dautenhahn (2004), "rather than relying on an inbuilt fixed repertoire of social behaviours, a robot should be able to learn and adapt to the social manners, routines and personal preferences of the people it is living with." These behaviours can play an important role in facilitating the interactions between users and social robots in several application domains and can ultimately lead to personalisation, which is a feature that is frequently mentioned to be relevant in social robots or agents that interact with users for extended periods of time (Castellano et al., 2008).

5. Where to Next?

We are entering an amazing new era for robotics, AI and Affective Computing because concrete emotional robots are being developed to a level in which they can be deployed in social settings. As new opportunities for affective robots appear, there will be new challenges and research problems. The affective computing community must embrace the challenges posed by the advancements in robotic platforms.

The community should consider these challenges that will shape the manner in which we build the emotional robots of the future:

- *Studies and more studies in real settings.* The deployment of emotional robots will necessarily indicate that tests should be conducted in valid social settings, in which emotional robots are placed in authentic situations. Such studies, although informed by Wizard of Oz (WOZ) tests, must be realised with autonomous systems as much as possible. This requirement poses a difficult challenge to the community, in terms of performance and reliability of the systems and in terms of the validity, believability and appropriateness of the emotional models adopted and expressions rendered.

- *New bodies for new behaviours and new situations.* Innovative new materials and mechanical engineering developments will facilitate novel embodiments of robotics. As the new embodiments become available to researchers in affective computing, we should combine them with new computational models of emotions, new forms of perceiving emotions and new implementations of expressions, and we should conduct studies of the outcomes.

- *Emotions are dynamic states, also for robots.* Emotional robots must be placed in real settings and must be capable of interacting with changing environments in a dynamic manner. We should use the dynamic features of emotions to create new models and perhaps, more neurobiologically inspired models that could provide additional adaptation to the changing environments.

- *Emotions are embodied.* Although robots have bodies, the majority of the work with emotions in robots treats emotions at a cognitive level. If we consider the embodiment aspects of emotional states, we can consider the various robotic embodiments associated with different emotions. Relying on human-like emotions is not required. With the development of new bodies, novel affective concepts may emerge as a synergistic endeavour between psychologists, computer scientists, neuroscientists, roboticists and cognitive scientists. Only in a multi-disciplinary way will we be able to design the next theories that will lead to the modelling of emotions in robots.

- *Can robots be our companions?* Companionship, attachment and long-term interaction with robots is definitely one of the greatest challenges to our community. This challenge involves new developments in learning, adaptation and memory, and extensive additional research is needed in this endeavour.

- *Affective symbiosis with robots.* The exploration of new types of links between humans and robots can be increasingly connected to humans and their physiological and affective states. These links range from physiological signals, BMI interfaces, and robotics platforms. Novel types of interdependence between robots and humans can be created and explored.

Acknowledgements

This research was supported by European Commission (EC) 7th FP project EMOTE under grant agreement no. 317923, and by national funds through FCT (Fundação para a Ciência e a Tecnologia), under project Pest - OE/EEI/LA0021/2011and the PIDDAC Program funds. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

Adolphs, R. A. (2002). Neural systems for recognizing emotions. *Current Opinion in Neurobiology, 12*, 169-177.

Álvarez, M., Galán, R., Matía, F., Rodríguez-Losada, D., & Jiménez, A. (2010). An emotional model for a guide robot. *Trans. Sys. Man Cyber. Part A* 40, 5, 982-992.

Arbib, M. A. & Fellous, J. M. (2004). Emotions: from brain to robot. *TRENDS in Cognitive Sciences*, 8, 554–561.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3567-3574.

Bartneck, C. & Lyons, M. J. (2009). Facial Expression Analysis, Modeling and Synthesis: Overcoming the Limitations of Artificial Intelligence with the Art of the Soluble. *Artificial Intelligence*, 33–53.

Bartneck, C. (2002). eMuu – An Embodied Emotional Character for the Ambient Intelligent Home Door.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Emotion* (April).

Beck, A., Cañamero, L. & Bard, K. A. (2010). Towards an Affect Space for Robots to Display Emotional Body Language. *IEEE International Symposium on Robots and Human Interactive Communications*, 491–496.

Bekele, E., Lahiri, U., Davidson, J., Warren, Z., Sarkar, N. (2011). Development of a novel robot-mediated adaptive response system for joint attention task for children with autism. In *RO-MAN, 2011 IEEE* (pp. 276-281). IEEE.

Bethel, C. L. (2009). Robots Without Faces: Non-verbal Social Human-Robot Interaction.

Bickmore, T., Picard, R. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* (TOCHI) 12 (2), 327.

Brave, S., Nass, C., Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. J. Hum.-Comput. Stud.*, 62 (2), 161–178.

Breazeal, C. (2003). Emotion and Sociable Humanoid Robots. *International Journal of Human-Computer Studies*, 59 (1-2) (July), 119–155.

Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1535), 3527.

Breazeal, C., Berlin, M., Brooks, A., Gray, J., Thomaz, A. L. (2006). Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54 (5), 385 – 393.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D. & Blumberg, B. (2004). Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots. *Artificial Life*.

Breemen, A. J. N. van (2004). Bringing Robots To Life: Applying Principles Of Animation To Robots. *CHI Workshop on Shaping Human-Robot Interaction*, 5-8.

Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P. W., Ho, S. (2008). Long-term affect sensitive and socially interactive companions. In *Proceedings of the 4th International Workshop on Human-Computer Conversation*.

Cramer, H., Goddijn, J.,Wielinga, B., Evers, V. (2010). Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *HRI'10: Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*. ACM, New York, NY, USA, 141–142.

Cutrona, C.E., Suhr, J.A. and MacFarlane., R. (1990). Interpersonal transactions and the psychological sense of support. *Personal relationships and social support*, 30-45.

Dautenhahn, K. (2004). Robots we like to live with?!-a developmental perspective on a personalized, life-long robot companion. In *13th IEEE International Workshop on Robot and Human Interactive Communication*, 17-22. IEEE.

Dautenhahn, K., Werry, I., Salter, T., Boekhorst, R. (2003). Towards Adaptive Autonomous Robots in Autism Therapy: Varieties of Interactions, *IEEE International Symposium on Computational Intelligence in Robotics andAutomation (CIRA'03)*, 577-582, Kobe, Japan.

Davis, M. H. (1994). Empathy: A social psychological approach. Westview Press.

Dias, J., & Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. *Progress in artificial intelligence*, Springer, 127-140.

Eisenberg, N., Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, 14, 131–149.

Eisenberg, N., Losoya, S. and Guthrie, I.K. (1997). Social cognition and prosocial development. *The development of social cognition*, 329-363.

Ekman, P. & Friesen, W. V. (1975). Unmasking the Face. Prentice-Hall.

Elfenbein, H., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203-235. doi: 10.1037//0033-2909.128.2.203

Feshbach, N. D. (1978). Studies of empathic behavior in children. *Progress in experimental personality research*, 8, 1–47.

Gielniak, M. J. & Thomaz, A. L. (2012). Enhancing Interaction Through Exaggerated Motion Synthesis. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*, 375.

Gonsior, B., Sosnowski, S., Buß, M., Wollherr, D., & Kühnlenz, K. (2012). An Emotional Adaption Approach to increase Helpfulness towards a Robot. *Proc. of IROS 2012*, Algarve, Portugal, IEEE.

Gratch, J., and Marsella, S. (2004). A domain-independent framework for modeling emotion, in *Cognitive Systems Research*, 5(4), 269-306.

Hanson, D., Baurmann, S., Riccio, T., Margolin, R., Docking, T., Tavares, M. & Carpenter, K. (2009). Zeno: a Cognitive Character. *AI Magazine*, 9–11.

Heberlein, A., & Atkinson, A. (2009). Neuroscientific evidence for simulation and shared substrates in emotion recognition: Beyond faces. *Emotion Review*, *1*(2), 162-177.

Hegel, F., Spexard, T., Vogt, T., Horstmann, G., Wrede, B. (2006). Playing a different imitation game: Interaction with an Empathic Android Robot. *In: Proc. 2006 IEEE-RAS International Conference on Humanoid Robots (Humanoids06)*. 56–61.

Hoffman, M. (2001). Empathy and moral development: Implications for caring and justice. Cambridge Univ Press.

Holroyd & A., Rich, C. (2012). Using the Behavior Markup Language for Human-robot Interaction. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*, 147–148.

Höök K. (2009). Affective loop experiences: designing for interactional embodiment. *Phil. Trans. R. Soc. B*, 364, 3585–3595.(doi:10.1098/rstb.2009.0202)

Kim, E. H., Kwak, S. S., Hyun, K. H., Kim, S. H. & Kwak, Y. K. (2009). Design and Development of an Emotional Interaction Robot, Mung. Advanced Robotics 23 (6), 767–784.

Kipp, M., Heloir, A., Schröder, M. & Gebhard, P. (2010). Realizing multimodal behavior: closing the gap between behavior planning and embodied agent presentation. In *Proceedings of the 10th international conference on Intelligent virtual agents (IVA'10)*, Springer-Verlag, Berlin, 57-63.

Koay, K.L., Syrdal, D.S., Walters, M.L., Dautenhahn, K. (2007). Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *16th IEEE International Symposium on Robot and Human interactive Communication, 2007. RO-MAN 2007*, 564-569. IEEE.

Kopp, S., Krenn, B., Marsella, S. & Marshall, A. N. (2006). Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. *Information Sciences*.

Krose, B. J. A., Porta, J. M., Breemen, A. J. N. Van, Crucq K., Nuttin, M. & Demester, E. (2003). Lino, the User-interface Robot. *First European Symposium on Ambient Intelligence*, 1–10.

Lasseter, J. (1987). Principles of Traditional Animation Applied to 3D Computer Animation. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques (SIGGRAPH'87)*, 35-44.

Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., Savetsila, S. (2012). Personalization in HRI: A longitudinal field experiment. In 7th ACM/IEEE International Conference Human-Robot Interaction (HRI'2012), 319-326.

Leite, I., Castellano, C., Pereira, A., Martinho, C., Paiva, A. (2012). Long-term Interactions with Empathic Robots: Evaluating Perceived Support in Children. *Proc. of 4th International Conference of Social Robotics*, Springer LNCS.

Leite, I., Martinho, C., Pereira, A. & Paiva, A. (2008). iCat: An Affective Game Buddy Based on Anticipatory Mechanisms. *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'2008)*, 1229–1232.

Leite, I., Martinho, C., Pereira, A., Paiva, A. (2009). As time goes by: long-term evaluation of social presence in robotic companions, *IEEE Int. Symp.on Robot and Human Interactive Communication, RO-MAN 2009*, 669-674.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., Paiva, A. (2013), The influence of empathy in human-robot relations, International Journal of Human-Computer Studies, 71(3), 250-260, ISSN 1071-5819, 10.1016/j.ijhcs.2012.09.005.

Liu, C., Conn, K., Sarkar, N., Stone, W. (2008). Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Trans. Robot*, 24, 883–896 (doi:10.1109/TRO.2008.2001362)

Ljungblad, S., Kotrbova, J., Jacobsson, M., Cramer, H. & Niechwiadowicz, K. (2012). Hospital Robot at Work: Something Alien or an Intelligent Colleague?. *Proc. of the ACM 2012 conference on Computer Supported Cooperative Work*.

McQuiggan, S. W., Lester, J. C. (2007). Modeling and evaluating empathy in embodied companion agents. *Int. J. Hum.-Comput. Stud.*, 65, 348–360.

Mead, R. & Mataric, M. J. (2010). Automated Caricature of Robot Expressions in Socially Assistive Human-Robot Interaction. *The 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI'10) Workshop on What Do Collaborations with the Arts Have to Say About HRI?*.

Moubayed, S. Al, Baklouti, M., Chetouani, M., Dutoit T., Mahdhaoui, A., Martin, J. C., Ondas, S., Pelachaud, C., Urbain, J. & Yilmaz, M. (2008). Multimodal Feedback from Robots and Agents in a Storytelling Experiment. *Digital Media*.

Niewiadomski, R., Obaid, M., Bevacqua, E., Looser, J., Le, Q. A. & Pelachaud, C. (2011). Cross-media Agent Platform. *Proceedings of the 16th International Conference on 3D Web Technology*, 11–19.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., Hall, L. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*. IEEE Computer Society, 194–201.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(04), 515-526.

Prendinger, H., Ishizuka, M. (2005). The Empathic Companion: a Character-Based Interface that Addresses Users' Affective States. *Applied Artificial Intelligence*, 19 (3-4), 267–285.

Quigley, M. & Gerkey, B. (2009). ROS: An Open-source Robot Operating System. http://pub1.willowgarage.com/~konolige/cs225B/docs/quigley-icra2009-ros.pdf

Rea, D., Young, J. E. & Irani, P. (2012). The Roomba Mood Ring: An Ambient-display Robot. In *Proceedings of the 7th annual ACM/IEEE international conference on Human-Robot Interaction (HRI'12)*, 217–218.

Reilly, W. S. N. & Simmons, R. (1996). Believable Social and Emotional Agents. *Science*.

Ribeiro, T. & Paiva, A. (2012). The Illusion of Robotic Life Principles and Practices of Animation for Robots. *In proceedings of the 7th annual ACM/IEEE international conference on Human-Robot Interaction (HRI'12)*.

Rich, C. , Ponsler, B., Holroyd, A., Sidner, C. L.. (2010). Recognizing engagement in human-robot interaction. *Proc. of the 5th ACM/IEEE international conference on Human-robot interaction*, 375–382, New York, NY, USA, ACM.

Riek, L. D., Paul, P. C., Robinson, P. (2010). When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3 (1-2), 99–108.

Russell, J. (1994). Is there universal recognition of emotion from facial expression - A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102-141.

Shindev, I., Sun, Y., Coovert, M., Pavlova, J. & Lee, T. (2012). Exploration of Intention Expression for Robots. *In proceedings of the 7th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*, 247–248.

Sidner, C.L., Kidd, C.D., Lee, C., Lesh, N. (2004). Where to look: a study of human-robot engagement. *Proc. of IUI 2004*, 78–84. ACM.

Singh, Ashish & Young, J. E. (2012). Animal-inspired Human-robot Interaction: A Robotic Tail for Communicating State. *In proceedings of the 7th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*, 237–238.

Takayama, L., Park, M., Dooley, D. & Ju, W. (2011). Expressing Thought: Improving Robot Readability with Animation Principles. *In proceedings of the 6th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'11)*, 69–76.

Tanaka, F., Noda, K., Sawada, T. & Fujita, M. (2004). Associated Emotion and Its Expression in an Entertainment Robot QRIO. *International Conference on Entertainment Computing*, 1–6.

Tapus, A., Ţăpuş, C., & Matarić, M. J. (2008). User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2), 169-183.

Thomas, F. & Johnston, O. (1995). The Illusion of Life: Disney Animation. Hyperion.

Torrey, C., Fussell, S., Kiesler, S. (2010). What robots could teach us about perspective-taking. In: Morsella, E. (Ed.), *Expressing oneself/Expressing ones self: A festschrift in honor of Robert M. Krauss*. Taylor and Francis, NY, 93–106.

Velásquez, J. D. (1998). When robots weep: emotional memories and decision-making. In *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence (AAAI '98/IAAI '98)*. American Association for Artificial Intelligence, Menlo Park, CA, USA, 70-75.

Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thórisson, K. R., Welbergen, H. van, der Wert, R. J. van, (2007). The Behavior Markup Language : Recent Developments and Challenges. *Artificial Intelligence,* 4722, 99–111.

Wistort, R. (2010). Only Robots on the Inside. *Interactions*.

Woods, Sarah, et al. (2005). Is this robot like me? Links between human and robot personality traits. In *5th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2005.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge engineering review*, 10(2), 115-152.

Zeng, Z., Pantic, M., Roisman, G. I., Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39–58.