Springer International Journal of Social Robotics
DOI: 10.1007/s12369-014-0262-y

A Recipe for Empathy

Integrating the mirror system, insula, somatosensory cortex and motherese

Angelica Lim · Hiroshi G. Okuno

Nov. 7, 2014

Abstract Could a robot feel authentic empathy? What exactly is empathy, and why do most humans have it? We present a model which suggests that empathy is an *emergent* behavior with four main elements: a mirror neuron system, somatosensory cortices, an insula, and infant-directed "baby talk" or motherese. To test our hypothesis, we implemented a robot called MEI (multimodal emotional intelligence) with these functions, and allowed it to interact with human caregivers using comfort and approval motherese, the first kinds of vocalizations heard by infants at 3 and 6 months of age. The robot synchronized in real-time to the humans through voice and movement dynamics, while training statistical models associated with its low level gut feeling ("flourishing" or "distress", based on battery or temperature). Experiments show that the post-interaction robot associates novel happy voices with physical flourishing 90% of the time, sad voices with distress 84% of the time. Our results also show that a robot trained with infant-directed "attention bids" can recognize adult fear voices. Importantly, this is the first emotion system to recognize adult emotional voices after training only with motherese, suggesting that this specific parental behavior may help build emotional intelligence.

Keywords robot empathy · emotional contagion based on SIRE model · MEI robot

Angelica Lim
Kyoto University, Kyoto
E-mail: angelica@kuis.kyoto-u.ac.jp

Hiroshi G. Okuno
Waseda University, Tokyo
E-mail: okuno@aoni.waseda.jp

1 Introduction

> "When dealing with people, remember you are not dealing with creatures of logic, but creatures of emotion." – Dale Carnegie

The year is 2030. A healthcare robot oversees an elderly patient named Linda at the local hospital. It is evening, and the robot is set to close the room at 9pm. Soaked by the rain, the patients daughter, Mary, knocks on the hospital room door. Mary has driven 50 kilometers from the airport, but a thunderstorm has delayed her arrival. She yearns to hold her mothers hand, because it has been 3 years since their last meeting. Linda's eyes light up as she sees her daughter through the hospital room window. But it is now 9:01pm. The healthcare robot locks the door with a loud thud. Crestfallen, the mom and daughter look pleadingly at the robot, but the rules are rules.

Robots do not share our capacity for empathy. It is easy to see why, in a 2012 survey, 60% of EU citizens stated that robots should be banned in the care of children, the elderly, or the disabled. Large majorities would also agree to ban robots from "human" areas such as education (34%), healthcare (27%) and leisure (20%). Yet, a new breed of service robots are advancing to our doorstep quickly, with the potential to change the lives of children and the elderly, able-bodied and disabled, students and more. For robots to be accepted in our daily lives as helpers, we must release robots from their pure, programmed logic and make them more emotional, more empathetic, to interact with humans on their own terms.

Empathy is the ability to experience the emotion of another person through implicit or explicit understanding of their position. Recent research from both evolution and neuroscience point to two different systems of empathy in humans, one based on emotional contagion, i.e., "I feel what you feel", and the other through advanced cognitive perspective-taking, i.e. "I understand what you feel." [21] [65]. According to Baron-Cohen et al., *empathy* not only allows us to experience an emotion triggered by someone else's emotion, but eventually also predict their behavior and understand their intentions [6]. But what are emotions?

Emotions have been defined as synchronized reactions involving multiple components, including subjective feeling, expression, physiological arousal, action tendency, and regulation [39]. For instance, in the above example, Mary might have reacted to the robot's rejection with a feeling of distress (subjective feeling), a downturned mouth and dejected voice pattern (expression), a heightened heart rate (physiological arousal), a plan to leave the hospital (action tendency), and an attempt not to cry in a public place (regulation). In this paper, we look at the first two components of emotion which emerge fast and involuntarily: internal feeling and external expression. *Feelings* are the private, mental experience of emotion, which we will further define in Section 3.1.3. *Moods* are different from emotions, in that they are lower in intensity, last longer than emotions (i.e. several hours to days) and are not necessarily in response to any clear event or object. Another useful term is *affect*, which has been defined as "an embodied reaction of pleasure or displeasure that references the goodness or badness of something and of arousal that references its urgency or importance" [15].

In the following sections, we describe the design and implementation of a robot with basic empathy. We first describe the benefits of robot empathy and discuss the challenges of its acceptance in Section 2. In Section 2 we will also give a brief overview of previous research in modeling robot emotions, and describe how our approach extends this work. We will then use observations in neuroscience in Section 3 and developmental science to build and support a new model of empathy. Finally, we will describe the implementation of the design on a Aldebaran NAO[1] robot, and present some experimental results with a robot that develops empathy in an human-robot caregiver loop in Section 4.1.

2 Background and Challenges

2.1 The Role of Empathy in Human-Robot Interaction

In the field of human-robot interaction, researchers have demonstrated the benefits of empathy in robot behavior design. Typically, a robot is trained to recognize the current emotion of a user, and then it mimics that emotion by changing its facial features, for example. Using the expressive robot EDDIE, Gonsior et al. showed that robots that mirror emotional states elicit more cooperative behavior from users [31]. Cramer et al. also showed that the iCat robot, when expressing accurate empathetic behaviors, was perceived as more dependable and trustworthy [17]. In terms of automatic systems, [46] [56] used a combination of facial expression recognition and rules to infer the human's emotional state, and showed that the robot was perceived more as a friend.

Empathy also works in the other direction, from human to robot. Riek et al. showed that the more human-like a robot is, the more humans empathize with them, which is interesting from a design point of view. They suggest a link with preferential treatment and societal in-group bias based on physical markers of similarity (e.g. skin color) [60]. Kwak and Kim [45] suggest that empathy – both in the human and in the robot – is important because it results in mutual emotional communication.

2.2 The Challenge of Authenticity

Despite the advantages that empathetic robots propose, one major challenge towards the acceptance of these empathetic robots is a perceived lack of authenticity. If a robot displays sadness in response to your grief, does it really *feel* sad? Does it matter? Among humans, expressing emotions that one does not really feel can be construed as fake, insincere, or even worse, manipulative. This has led to strong concerns, for instance by Turkle, that elderly care or child companion robots would be deceiving their owners by providing inauthentic affective relations, because robots do not have real "feelings" [70].

2.3 Towards Authentic Emotional Robotics

What exactly does it mean to be "authentic"? Authentic is defined as 1) "conforms to an original so as to *reproduce essential features* (e.g., an authentic reproduction of a 16th century castle)" and 2) "*made or done in the same way* as an original (e.g., authentic Mexican cuisine)"[2]. In this paper, we will suppose that we are referring to robot authenticity relative to the original—the human.

In order to understand how the challenge of empathetic authenticity could be addressed, let us review

[1] http://www.aldebaran-robotics.com

[2] www.merriam-webster.com/dictionary/authentic

some previous approaches towards building robot emotion systems. A first approach is to equip a robot with a fixed emotion architecture (*essential features*) inspired by theories and models of biological systems. A second approach is to create a robot software architecture that, in addition to mirroring human (i.e. brain) architecture, continually learns from its environment as a child does (*made or done in the same way as the original*). This paper situates itself in the latter, in the relatively new field of developmental robotics.

2.3.1 Modeling Components from Biological Systems

Many robot emotion systems exist (e.g. [71] [74]), but perhaps the most well-known emotional robot system to date is MIT's expressive robot Kismet from the late 1990's. Breazeal et al. [11] modeled Kismet's motivational system as an interconnected network of modules mirroring those found in biology, including modules such as drives and situational appraisals. For example, when the robot achieved one of its goals, the internal elicitors pulled the robot's internal emotional state closer to happiness, and the robot expressed itself accordingly. Kismet's emotional architecture was based on a system of needs and drives and models of emotion from psychologists (e.g. Mehrabian's PAD emotion representation [50]). Furthermore, while traditional emotion models used purely animal theories (e.g. fear/avoidance functions in a rat), Kismet aimed to handle human-specific behaviors such as raising eyebrows and smiling. Kismet could both perceive human emotions and express them, with underlying software modules that mirrored human mental architecture, and this integration was a significant step towards the goal of building robots with human-like emotions.

2.3.2 Developmental Robotics

Kismet was also one of the first projects which placed the human and robot in caregiver and infant roles. For instance, the robot responded to infant-directed speech, or baby talk, which will be discussed later in this paper. These unique interactions with Kismet pioneered the beginning of a new field of robotics and cognitive modeling based on infant psychology, called developmental robotics.

The essential goal of developmental robotics (also known as epigenetic robotics) is to study developmental mechanisms and architectures, for a robot to learn as a child does. It involves formalizing, validating and extending models from neuroscience, developmental psychology, and evolutionary biology, specifically by attempting to implement the models in robots. At the same time, these formalized models are expected to feedback into existing theories, or produce novel theories about human development. For instance, roboticists such as Nagai and Rohlfing created models to describe the use of motion as an early mechanism for teaching [53], and Ishihara has studied language development using this paradigm [37]. A comprehensive review of the field is available in [3].

One notable early work in emotional developmental robotics is done by Kozima et al., who explore the idea of robot empathy built like that of children's [44]. In that work's final section, Kozima provides an outlook for emotional robots, including the development of a robot "mirror system", a concept we will explore later. These interesting ideas were not yet implemented in a robot at the time. As a general rule, it is thought to be useful to implement models in robots when elaborating a model, because the implementation step itself may expose holes in understanding within the model.

Despite the engineering nature of robot implementation, we should make it clear the primary goal of developmental roboticists is the creation of a better, more formal models itself. As Scassellati, a pioneer of the developmental robotics field writes, "Developmental robotics is not about creating a system with peak performance at an engineering task. At first glance, it seems that those who employ computational models to describe developmental phenomena belong in the same camp as those who want to build better computational systems using inspiration from developmental biology. In both cases, the goal is the creation of a computational framework, simulation, or mathematical formulation. [...] However, the aims of these two groups could not be more dissimilar: one group uses biological themes to develop better engineering; the other uses computational techniques to formulate better descriptions of biological development." [63].

As such, we take the latter, developmental robotics approach towards empathy: by creating a better, formalized model of empathy, the closer we may approach to human-like processes, and perhaps, authenticity.

2.4 Our Approach

Since the time of Kismet, significant progress has been made in functional magnetic resonance imaging (fMRI) technology towards understanding the functional components in the brain. Therefore, we extend previous emotion models in the following ways: 1) We use the latest neuroscientific evidence to build our theories on empathy, 2) We take a developmental robotics approach, assuming that empathy, just like other skills such as sensory-motor associations, could be developed through early caregiver interaction, and 3) We address the concept of *feeling*, which has not been addressed by any other robot emotion system to-date.

In the next section, we will present our new model along with evidence supporting it. Firstly, we will identify the relevant neural architectures of empathy, which should help us reproduce *essential features* in the brain for empathy. In particular, we expect that neuroscientific evidence should help us address visible motor behavior as well as the essential but thorny question of inner feelings. Secondly, we will model the *development* process. How do human infants develop into empathetic beings? We did not know the precise answer to this, so a major contribution of this paper is to propose a concrete developmental explanation.

3 Empathy Systems in Humans

Let's now examine three major concepts involved human empathy systems:

- **Affective empathy** through emotional contagion,
- **Physical, gut feelings**, a way to ground emotional expressions,
- **Infant development**, to construct associations between expression and feeling.

Recent neurological studies have provided evidence that humans have two independent systems for empathy in the brain: affective empathy and cognitive empathy [65]. *Affective empathy* is thought to rely on a basic emotional contagion system, and *cognitive empathy* on a cognitive, perspective-taking system. For example, in the introductory scenario, there are two ways that a human nurse may have been affected by Linda and Mary's predicament. Mary could have explained her situation about the rain, the accident, and the long-awaited reunion. Perhaps the nurse would have imagined herself in their situation, and opened the door. This is cognitive empathy. But the *way* Mary asked is also important. If Mary and Linda had plead, with sorrowful faces and voices, an empathetic nurse would also be hard-pressed to keep the door locked. This second type is called affective empathy and is based on a mechanism that is "automatic, unintentional, uncontrollable, and inaccessible to awareness," called emotional contagion [33].

Robots may benefit from an affective empathy system. Simon Baron-Cohen has argued that narcissistic, borderline and psychopathic personalities may have intact *cognitive* empathy (e.g., a psychopath may recognize that their victim is in pain) but lack *affective* empathy (e.g., not actually feel pain themselves while torturing a victim) [7]. In a dictator game experiment, altruistic sharing behavior was found to be related to affective empathy, whereas cognitive empathy was not [24]. Since affective empathy appears to have certain desirable characteristics which cognitive empathy does not, we will focus on the former in this paper.

One notable element of affective empathy its underlying mechanism called *emotional contagion*, defined as "the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally" [33]. This convergence may even surpass emotions. Preston and De Waal have contrasted emotional contagion with other kinds of empathy: in cognitive empathy and sympathy, the separation between the self and other is retained, but in emotional contagion, there is no *self-other* distinction [58]. In other words, during emotional contagion, the "self-other" distinction becomes blurred. How does emotional contagion work?

3.1 Emotional contagion

Most studies focus on emotional contagion through facial expressions [33], but it has been found in other modalities such as voice. Neumann and Strack have found evidence of emotional contagion from speech [54]. In a covert text-comprehension task, participants listened to a text read in a slightly happy, neutral, or slightly sad tone of voice. They found that those who listened to the happy voice rated their mood significantly better than those who listened to the sad voice:

> "Even though participants did not have the goal of sharing the feelings of the target person, exposure to the emotional expression promoted a congruent mood in the listener." [54]

Emotional contagion is also used by recording artists and film score directors to induce emotion through music [39]. In a 2008 study, Juslin et al. provided participants with handheld devices to record their musical emotional experiences at random points throughout a 2-week period [38]. In 64% of the episodes when the participants were listening to music, they reported that the music influenced their feelings, and over 573 musical episodes, *emotional contagion* was reported most often as the reason (32%).

To summarize, emotional contagion has two parts:

- Automatic mimicry and synchronization of facial expressions, vocalizations, postures and movements with those of another person's
- Emotional convergence

To design a solution based on these requirements, we take inspiration from neurological studies. A 2009 review provides a broad summary of what is known about empathy and the brain [35]:

> "A large-scale network for empathy is composed of the mirror neuron system, the insula, and the limbic system."

3.1.1 Mirror neuron system

The mirror neuron system has been suggested to be a major mechanism for empathy [36, 55]. Mirror neurons (also known as the mirror mechanism) are those in the motor areas of the human brain that fire both during action execution and also action observation [35]. The first study in 1992 showed that the neurons in the premotor cortex of a macaque monkey grasping food were also found to be active when observing a human grasping food [22]. More recent studies show more direct links between mirror neurons and emotion. A functional MRI experiment found a mirror neuron network that was active both when observing a facial emotion and imitating the emotion [14]. Further, [55] has found evidence that affective empathy engages the mirror neuron system more than cognitive empathy.

Mirror neurons have been proposed as a critical step towards the simulation of the mental states of others, by mapping sensory input to internal representations. For example, mirror neurons provide precise visual to motor mapping in the studies described earlier, where simple visual observation of an action incited premotor activity in the brain [61, 36]. Similar results have also been found for auditory input: neurons in the monkey premotor cortex discharge both when it performs an action and when it hears the related sound [43]. In a neurological music study, simply listening to music activated brain areas related to premotor representations for vocal sound production (though no singing was observed in participants) [42]. This internal representation via the mirror neuron system is similar to what Damasio called an "as-if-body-loop" mechanism for emotion:

> "The brain momentarily creates a set of body maps that does not correspond exactly to the current reality of the body." [19]

In summary, mirror neurons provide a "sensory-motor gateway for forming an internal representation of an observed person's state" [23]. Therefore, our robot implementation should have an *internal representation* of the other's body state, not a simple one-to-one imitative mapping. This insight can help us respond to our emotional contagion requirement of "automatic mimicry and synchronization of facial expressions, vocalizations, etc.", and its implementation is further elaborated in Section 4.1.

In addition to motor mimicry, there should also be an "emotional convergence" for emotional contagion: that is, a sharing of an emotional feeling:

> "We understand what others feel by a mechanism of *action representation* that allows empathy and *modulates our emotional content*. The *insula* plays a fundamental role in this mechanism." [14]

3.1.2 The insula

A part of the brain called the *insula* has been suggested to be at the crux of the association between action representation and emotion [14]. All mammals have an insula that reads their body condition, by way of visceral and interoceptive sensors (sensing heat, cold, pain, taste, muscle ache) sending information to the insula [12]. In humans, great apes, whales, and elephants, it contains a special type of cell called the Von Economo Neuron, which are large and are hypothesized to help channel neural signals from deep within the cortex to relatively far parts of the brain [16].

The insula has been associated with many behaviors: drug cravings, feeling pain, maternal love, empathizing with others, seeing disgust on a face, and listening to music [16, 52]. It thought to be where bad taste or smell are transformed into disgust [2, 13], or a sensual touch into pleasure [51]. It is active when a mother hears a crying baby [40], or when a person looks at a happy face [57]. It is active both when empathizing for others' pain [67], and when directly feeling pain [8].

How exactly does the insula work? Although its exact mechanism is still not clear, Damasio and others have proposed that the human insula plays a role in mapping visceral states that are associated with emotional experience, giving rise to conscious feelings [19, 16, 67]. Therefore, an integral part of an empathetic robot system appears to be an *artificial insula*: a module that associates an emotional experience to a robot's own physical, bodily state.

A final component related to empathy is the feeling itself. Insight can come from brain lesion studies such as [1] by Ralph Adolphs. In this study, patients looked at photographs of people expressing an emotion and were asked to guess the person's state of mind by "placing himself in the person's shoes". The most interesting result was those with impaired performance in this task: those with damage in the insula and somatosensory cortices.

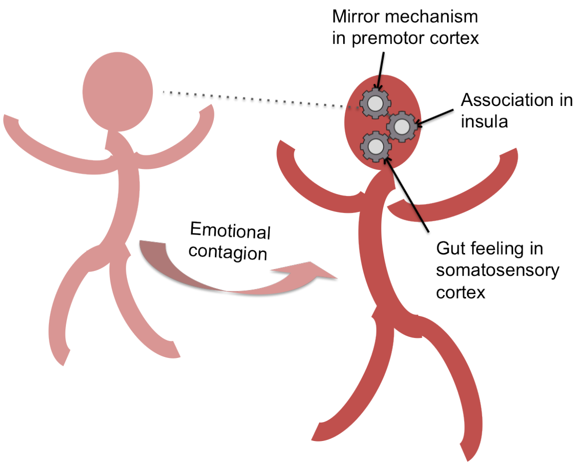

Fig. 1 Proposed mechanisms necessary for emotional contagion.

3.1.3 Physical, gut feelings

In the brain, the group of somatosensory cortices (from the Greek root *soma*, meaning body) are responsible for sensing the body's internal state including viscera (e.g. internal organs such as heart, stomach, lungs) and joint position, as well as the external senses of touch, temperature and pain. Damage to these areas can have high level, emotion-related repercussions: brain patients that have damage to the cortices in charge of processing signals from the body do not show normal signs of despair or panic ([18], pg. 63).

The somatosensory cortices are important, because they form the basis of one of the most well-known theories on *feelings*, called the Somatic Marker Hypothesis [18]:

> "Feelings are [...] first and foremost about the body, that they offer us the cognition of our *visceral and musculoskeletal state... By [...] juxtaposition, body images give to other images a* quality *of goodness or badness, of pleasure or pain."*

In short, Damasio's Somatic Marker Hypothesis suggests that feelings are an association of stimuli to visceral (and musculoskeletal) pleasure or pain. He also suggests that "the critical, formative set of stimuli to somatic pairings is, no doubt, acquired in childhood and adolescents." ([18] p.179)

3.2 The triad: mirror system, association, and gut feeling

Based on these findings in neuroscience, we offer an architecture for robot empathy in Fig. 1, and a summary

Human behavior	Mechanism	Related brain areas
Internal & external mimicry	Mirror mechanism (& visible synchrony)	Premotor cortex (Motor cortex)
Emotion association	Stimuli-viscera association	Insular cortex
Internal feeling	Physical, gut feeling	Somatosensory cortices

Fig. 2 Proposed essential features for an emotional contagion system, with related locations in the brain.

of the features in Fig. 2. A robot with empathy should model at least these three areas of the brain: a) mirror neurons in the premotor cortex, b) the insula c) and the somatosensory cortex. These correspond to three functional modules in a robot system:

- **A mirror system**: represents the action of another human and can explain eventual motor imitation (Premotor cortex)
- **An associative module**: associates an action representation with a physical feeling of pleasure or pain (Insula)
- **A gut feeling module**: a module for visceral and musculoskeletal pleasure or pain, e.g. battery level or motor heat (Somatosensory cortex)

Emotional contagion would thus work as follows. The robot would first observe another person performing an emotive action (such as a sad voice or movement), and create an action representation in the mind. This "as-if" body representation would trigger two things: 1) a visible motor mirroring response, and 2) a learned association between the action and a physical, gut feeling of "goodness" or "badness" (for example, sufficient or insufficient battery level in a robot). It is through this associative mapping that seeing another person's pain would cause an automatic, visceral pain in oneself, for instance. These two reactions correspond to our requirements for emotional contagion: 1) automatic mimicry and synchronization of facial expressions, vocalizations, postures and movements with those of another person's, and 2) emotional convergence.

We should also note a glaring omission in our survey: a review of emotion literature on the brain should include mentions of other parts of the limbic system – especially the amygdala. We consider the integration of the amygdala's functions as future work for two reasons. First, while the amygdala's role in autonomic fear responses related to survival are clear [20], the insula predominates in studies related to empathy for pain, for example [34]. Further, computational emotion models focusing on the fear response have been well established, for instance by Fellous and LeDoux [26]. While not diminishing the amygdala's role in emotions, we strive here to provide a complimentary perspective that can extend to empathetic emotions and be implemented in an embodied robot.

3.3 Developing empathy in childhood

A final puzzle piece for our system is the association that happens in the insula. How does an action representation trigger the appropriate visceral response? For example, how would a robot encountering a happy person retrieve a pleasant, and not painful feeling? Damasio suggested that "the critical, formative set of stimuli to somatic pairings is, no doubt, acquired in childhood and adolescents."([18] p.179). Indeed, dramatic development of emotional intelligence occurs even before infants reach the age of 1 year old:

> "Although the development of emotion perception extends beyond infancy–perhaps throughout the lifespan–[...] *dramatic changes in emotion perception competencies [...] are observed over this period of development [during the first year of life].* Furthermore, it may be that infants reared in situations with impoverished affective expression information, such as those, for example, from caregivers with clinical depression, or in contexts where actions and expressions are discrepant, may be particularly influenced in their comprehension of expressions." [72]

Our idea for developing this association lies in an empathetic mirroring that occurs between caregivers and their infants. In developmental psychology, [72] and [32] have suggested a kind of associative learning between the affective voice and face during infancy. That is, the infant receives multiple modes of emotional input (such as a smile and a happy voice) simultaneously, causing an association between these visual and auditory streams. Further, a parent may produce emotional mirroring, such as facial and vocal signals of sadness upon seeing their child cry. In developmental robotics, this "intuitive parenting" paradigm has been proposed for associating a viewed emotional face with a robot's own emotional face [73, 10]. Although these approaches "ground" the perceived facial expressions with a robot's own facial expression, no developmental robot research goes as far as to explain how a feeling may be associated. For instance, Breazeal specifically states that Kismet does not address the concept of feelings [11].

We propose to develop a robot's empathy based on an emotional interaction called motherese. Recent research has suggested this universal mechanism, also called infant-directed (ID) speech, is at the crossroads for developing emotions, cognition and language in humans [62]:

> "It has been shown that infants are attracted by and attend to motherese, which is characterized by more exaggerated intonation and higher pick than adult-to-adult speech. Concurrent with the exaggerated speech of motherese, there are probably exaggerated facial displays, allowing infants to explore the particular aspects of the face (e.g., exaggerated mouth and brow movement). [...] Child-centered displays may serve as opportunities for learning about affective events." [68]

For example, expressions of approval such as 'Good!' or 'Clever girl!' are typically spoken using exaggerated rise-fall pitch contours [and] expressions of prohibition or warning such as 'No!' or 'Dont touch that!' are spoken with low pitch and high intensity [27]. Most interestingly, motherese appears to have correlates with emotional speech among adults:

> "Acoustic analyses showed few differences between the infant-directed and adult-directed [emotional] samples, but robust differences across the emotions. [69]"

To summarize, a motherese-like, empathetic interaction between a caregiver and a robot can provide input that is both a) emotional and b) consistent with the robot's own internal state. For example, when an infant cries due to hunger or pain, a mother may soothe her with an empathetic voice; when an infant is in a flourishing state, the mother provides positive vocal sounds and smiles. This empathetic interaction is how we propose to develop a robot's association between an action (such as a voice or movement) and a visceral, gut feeling.

4 A robot that develops empathy through interaction

In this section, we describe a robot implementation of the emotional contagion architecture presented in Section 3, along with some experimental results.

4.1 Design of an empathetic human-robot feedback loop

We have covered some of the emotional characteristics of motherese (ID speech) itself, but what are the characteristics of a motherese-like interaction? According to Gleason [30], it is not a one-way communication – it requires two active participants: when a caregiver speaks to an infant, the infant's *reactions* shape the interaction. In fact, Fernald has shown that mothers

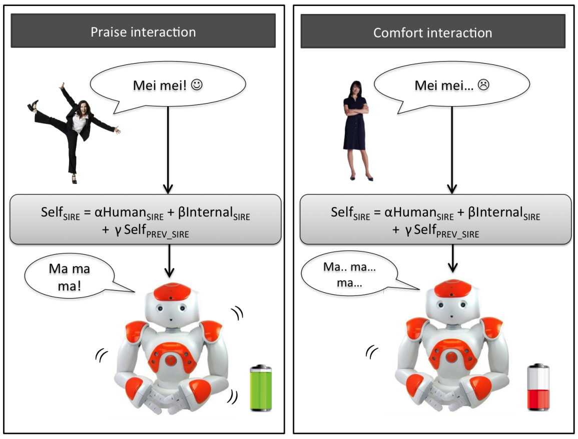

Fig. 3 The robot synchronizes to the caregiver by extracting SIRE parameters from the human's voice and reproducing them in speech and gesture. The output is a combination of observed $Human_{SIRE}$ and $Internal_{SIRE}$ (a SIRE it has learned to associate with its current internal state), dampened through time.

cannot reliably produce motherese it in front of a microphone [29]. *Turn-taking* is also observed early vocalizations between a mother and infant [59], which has been described as "mutual entrainment between mothers an infants during early social interactions" [9]. According to some studies, this may even involve correlations of melody types [62]. Inspired by these findings from infant psychology, we design a robot system that allows a back-and-forth interaction, where the robot takes the place of the infant.

In the present implementation, we will focus on empathetic vocal and motor movements. This is because we shall use a NAO robot, which does not have a moveable face capable of making facial expressions. We use the SIRE paradigm [48], in which vocal utterances and motor movements are abstracted into their dynamic characteristics: Speed, Intensity, irRegularity, and Extent (SIRE). This abstraction is chosen because dynamics such as SIRE have been shown to underlie emotion across multiple modalities [47] and even across cultures [66]. For example, slow, non-intense and small vocalizations and movements are perceived as sad.

How do we design an empathetic feedback loop? The scheme for human-robot synchronization is shown in Figure 3. The robot entrains to the dynamics of the caregiver's utterance by extracting speed, intensity, irregularity and extent (SIRE) from the human's vocalizations, $Human_{SIRE}$, and reproducing them in gesture and speech. However, it is more than simple mimicry—we also integrate internal state of the robot. Consider that a hungry infant could be soothed momentarily by a toy, but inevitably it would express distress until it was fed. Similarly, in the robot expresses a combination of a) what it sees and b) its own internal emotional state. $Internal_{SIRE}$ represents the dynamics (SIRE 4-tuplet) it has learned to associate with its current internal state until that moment. In addition, the effects are dampened through time by deviating from the previous SIRE expression.

4.2 Robot gut feeling

At birth, the human infant is equipped with the most innate of emotional expressions: crying. Indeed, at this point in development, the infant is in one of two physical states: homeostasis, or not (e.g., extreme heat or cold, empty stomach). This in-built distress signal of crying alerts the caregiver to a lack of homeostasis.

Inspired by newborns, we define a robot's most basic level of physical "feeling" based on these two states. Ortony calls the most innate emotional level the *reactive level* which "assigns along two output dimensions, one of which we call "positive" and the other "negative"" [25]. Similarly, Damasio defines feelings as "the expression of human flourishing or human distress, as they occur in mind and body."[19]. Therefore, we define two "innate" physical, gut feeling states for our robot, which we will call *flourishing* and *distress*. In this paper, these two states are represented symbolically, but in future work they should be tied to a robot's physical state, for instance *flourishing* corresponding to full battery and CPU/motor temperatures within working limits, and *distress* corresponding to a near-empty battery and/or hot motors.

This physical, gut feeling $F = (flourishing, distress)$ has 2 important functions. The first is to cause a distress signal to be emitted when the robot is in *distress* state. The second is to serve as a switch (cf. the diamond symbol in Figure 4) for two sub-functions: a) storing information into long-term memory based on the value of F and b) retrieving information from long-term memory based on the value of F. For example, if the robot is in a *flourishing* state when a caregiver is smiling and speaking to it in a happy voice, the system will store this "happy voice" information in the $F = flourishing$ state (Figure 4). If the robot is in a *distress* state and the caregiver tries to comfort it with a soothing tone, the robot will store this "comforting" vocal information into the $F = distress$ state. At the same time, if the robot is in a flourishing state, it will sample from the trained GMM to produce an internal SIRE value, and so on.

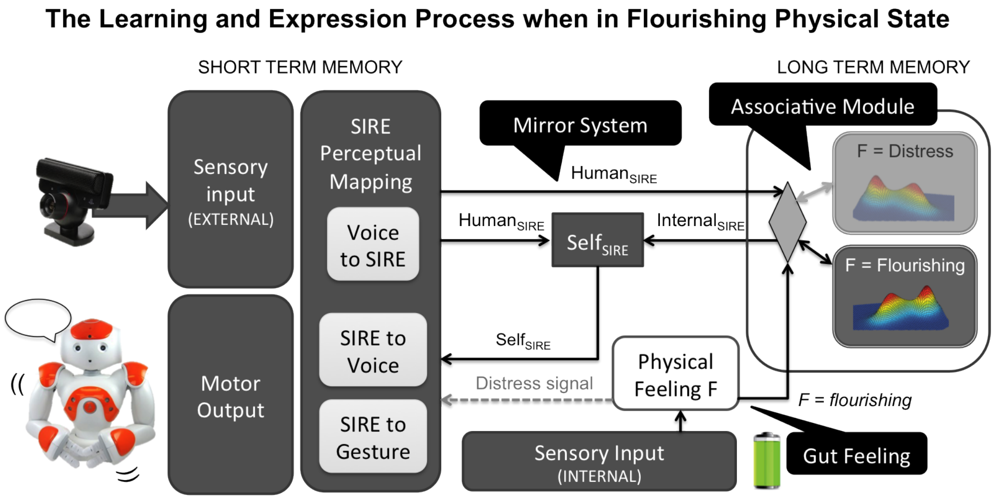

Fig. 4 An overview of the system when the robot is in a flourishing state. The diamond symbol activates flourishing or distress based on physical feeling F. The distress signal is active only when the robot is in a distressed state (e.g. low battery.)

4.3 Using SIRE Gaussian Mixtures Model as emotional long-term memory

How is association between SIRE representations and gut feelings stored into long-term memory? Here, we use a statistical learning paradigm called SIRE Gaussian Mixture Model (SIRE GMM), which we proposed in [49]. Each GMM in SIRE space represents a class C of *flourishing* and *distress*. We define an m-mixture Gaussian in 4D SIRE space,

$$SIRE\_Emotion_c(X_c) = \sum_{k=1}^{m} \pi_k \mathcal{N}(X_c|\mu_k, \sigma_k) \quad (1)$$

where X_c is a vector of SIRE tuples corresponding to the class C, and m is the optimal number of components to minimize the Bayesian Information Criterion (BIC) over X_c [64].

Put simply, the essence of an emotional expression is represented by four numbers, corresponding to speed, intensity, irregularity and extent (SIRE). We create two statistical models of these SIRE values, which are linked to the physical states of flourishing and distress, respectively. Together, we refer to the collection of these statistical representations (SIRE GMMs) as Multimodal Emotional Intelligence, or MEI.

Learning To illustrate, let us assume that the robot is in a flourishing (i.e., full-battery) state. A human, following the intuitive parenting paradigm described earlier, may begin to speak to the robot in a happy way. In this case, since the robot's physical feeling $F = flourishing$, the human's observed SIRE values $Human_{SIRE}$ will be added to the training data for the *flourishing* SIRE GMM.

Expressing The robot's expressed gesture and vocal dynamics $Self_{SIRE}$ depend both on the human's SIRE dynamics and the robot's $Internal_{SIRE}$. The value of $Internal_{SIRE}$ is produced by sampling from the statistical distribution corresponding to the current physical feeling F. For example, when the robot's physical feeling $F = distress$ (i.e., low battery), we sample from the *distress* SIRE GMM.

As an intermediate step, we can imagine the robot's voice and movements being modified using $Self_{SIRE}$, a vector of four values on $[0, 1]$, where

$$Self_{SIRE} = \alpha Human_{SIRE} + \beta Internal_{SIRE}, \quad (2)$$

and $\alpha + \beta = 1, 0 \leq \alpha, \beta \leq 1$. This equation is further refined in the following sections.

Empathy: The ratio of imitation and internal state. How do we decide the values of α and β? Consider that if $\alpha = 1$, the robot shows perfect mimicry. If $\beta = 1$, then the human in front of the robot has no immediate effect on the robot's expressions, and the robot simply expresses based on its internal physical feeling. These values can be considered a kind of empathy setting for the robot. For example, if the human is expressing sadness, then a robot with $\alpha = 1$ would immediately portray similar sad vocal and gestural dynamics (high empathy). On the other hand, a robot with $\beta = 1$ and a full-battery state would simply convey what it has learned to associate with it's own flourishing physical feeling, ignoring the sad expressions of the human (low empathy). In our experiments, we generally set α and β to be equal, but it would be interesting in future work to test the impression of the robot when changing these parameters.

Entrainment Entrainment is a term used to designate synchronizing with and adapting to the interaction partner ([5], p. 134). As shown in Figure 3, we deduce

Voice feature	Parameter	Gesture feature
Speech rate	Speed	Arm velocity
Change in volume	Intensity	Arm acceleration
High-frequency energy	irRegularity	Inter-arm phase shift
Pitch range	Extent	Gesture extent

Table 1 Low-level feature to SIRE mappings.

our current SIRE state based on the previous emotional state, to designate a temporal entrainment between the caregiver and the robot. Here is our final equation for empathetic entrainment:

$$Self_{SIRE} = \alpha Human_{SIRE} + \beta Internal_{SIRE} + \gamma Self_{PREV\_SIRE}, \quad (3)$$

and $\alpha + \beta + \gamma = 1, 0 \leq \alpha, \beta, \gamma \leq 1$. In our experiments, we set $\alpha = \beta = \gamma = 1/3$, and future work should test other configurations for these parameters.

4.4 Real-time implementation of SIRE audio processing

We implemented the system described in Figures 3 and 4 using the Aldebaran NAO robot and HARK (HRI-JP Audition for Robots with Kyoto University)[3] real-time robot audition system. A Playstation Eye was used as a microphone input, to avoid mixture with the robot's speaker output. The speech recognition system Julius, trained with an English acoustic model, was used to detect the words spoken by the user, in order to detect the speed of the vocalizations. The mappings for SIRE voice and gesture, and SIRE GMM learning mechanism remained the same as defined in [48], shown in Table 1. A demonstration video is available at this address: `http://youtu.be/f9F8FSVhBwM`.

5 Experiment 1: Online collection of motherese

5.1 Purpose

The goal of this experiment was to test the system in an online manner with naive experiments, and collect the SIRE dynamics of their utterances for analysis.

5.2 Materials and procedure

We used the robot motherese system to collect infant-directed samples of praise, comfort, prohibition and attention from human participants. These correspond to the four categories of motherese as defined in [27].

[3] http://www.hark.jp

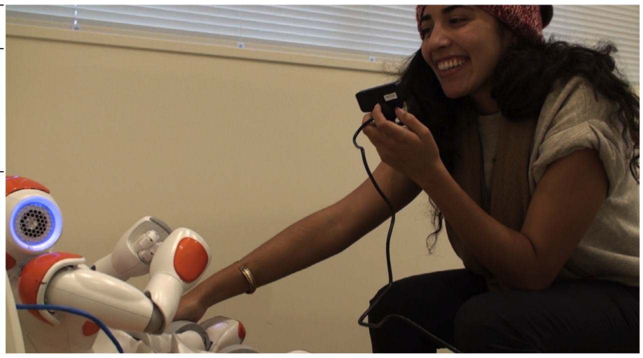

Fig. 5 A participant interacts with the robot by speaking into a microphone.

We recruited 6 fluent English speakers from Western countries (3 female, 3 male, mean age=29.8 years, std=3.9): 3 from USA, 1 Australia, 1 Madagascar, 1 Chile. The participants were first introduced to the robot through a written introduction, and told that the robot's name was "Mei Mei". The robot was introduced as "young and continuously learning", and that the participants were the robot's caregiver for the duration of the experiment. The participants were also told that the robot, because it is young, could not understand the content of their words, only the way in which they say it.

The participants were then instructed to interact with the robot in four different motherese situations, by speaking into the microphone and saying the robot's name, as in Figure 5.

1. **Attention**: Get Mei Mei's attention by saying her name.
2. **Prohibition**: Mei Mei is crying. You will try two different ways to stop her from crying. First, you will prohibit her from crying by saying "Mei Mei".
3. **Comfort**: Mei Mei is crying again. Your goal is to soothe and comfort her by saying "Mei Mei".
4. **Praise**: Your goal is to praise Mei Mei because she is no longer crying, and make her feel that she is loved.

For the purposes of controlling the experiment, the physical feeling F was set manually to correspond to the situation. In Situation 1 and 4, $F = flourishing$, and in Situation 2 and 3, $F = distress$. In all four situations, the robot gestured to convey its affect using $Self_{SIRE}$. In the 2nd and 3rd situation, to simulate a distress signal, the robot also vocalized. The "cry" was produced by repetition of the syllables *"ma ma ma"*. This vocalization was also subject to $Self_{SIRE}$. In the distress situations (2 and 3), $Self_{SIRE}$ was initially to [0.9, 0.9, 0.9, 0.9]. In the flourishing situations (1 and 4), $Self_{SIRE}$ was initialized to [0.1, 0.1, 0.1, 0.1]. The in-

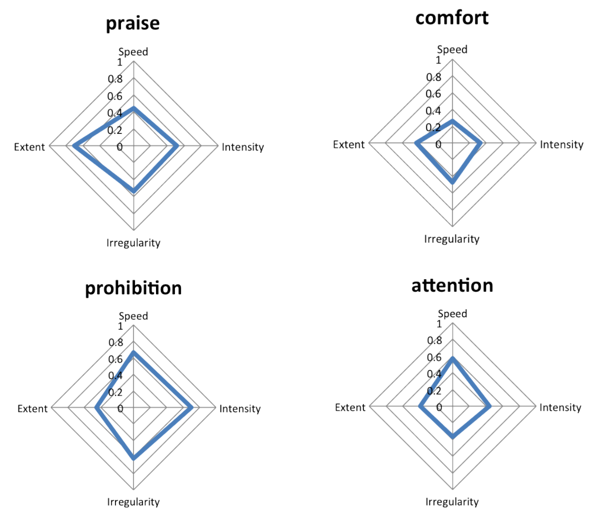

Fig. 6 Plotting the SIRE means of 1-mixture GMMs trained in each condition.

Fig. 7 Visual input accompanying the different kinds of "Mei Mei" vocalizations: praise (top left), comfort (top right), prohibition (bottom left), attention (bottom right). Images captured from robot's camera during each condition, mid-utterance.

teractions, of course, modified the robot's internal state continuously, based on Equations 4 and 5. All initial settings were identical for each participant.

The interactions were recorded with a standard video camera, as well as through the robot's own front-facing camera.

5.3 Results

The interactions resulted in 510 motherese utterances in total (128 praise, 114 comfort, 123 prohibition, 145 attention). We plot the means of the resulting GMMs in Figure 6. These correspond to the average dynamics presented to the robot during each condition. We also show example captures from the robot's camera during vocalizations in each condition in Figure 7. The variation across motherese vocalizations and facial expressions gives qualitative weight to our hypothesis that our robot system can elicit motherese and facial expression that are differentiable across interaction types (praise, comfort, prohibition and attention).

6 Experiment 2: Training with motherese, testing with emotional voice

6.1 Purpose

The goal of this experiment was to test if motherese could be an adequate training mechanism for learning emotion dynamics. We do this by performing two analyses.

First, we test whether the robot could associate happy voices with a physical flourishing state, and sad voices with a physical distress state. Indeed, if a robot listens to a sad voice and associates it with its own experience of distress, this could provide evidence that our approach provides a means for robot empathy. When the participants praised and comforted the robot, they spoke to the robot while it was in a physical state of flourishing and distress, respectively. We therefore train two SIRE GMM models on the caregiver input during these interactions and check how they respond to happy and sad voices.

Secondly, we hypothesize that all the utterances in motherese (praise, comfort, prohibition and attention) have correlates with adult-directed expressions of emotions. If that is true, then our robot could learn to associate emotional voices to situations in which it heard similar motherese utterances.

6.2 Materials and procedure

The utterances captured in Experiment 1 were used to train two different MEI.

- The first MEI contained two GMMs, one for each of the conditions *flourishing* and *distress*. The flourishing GMM was trained with samples collected during the praise condition, and the distress GMM was trained with samples from the comfort condition.
- The second MEI contained four GMMs, one for each of the conditions praise, comfort, prohibition and attention.

Next, the emotional voice samples from an emotional voice database called the Berlin Database of Emotional Speech[4], which is a database of emotional speech

[4] http://pascal.kgw.tu-berlin.de/emodb/

Detected Input	**Flourishing** # associations	**Distress** # associations
Happiness	**64** (90%)	7 (10%)
Sadness	10 (16%)	**52** (84%)

Table 2 Emotional voice association detection rates on a motherese-trained model

Detected Input	**PRA** (%)	**COM** (%)	**PRO** (%)	**ATT** (%)	*p-value*
Happiness	**54**	10	24	13	0.0001
Sadness	0	**65**	2	34	0.0001
Anger	**47**	8	38	7	0.0001
Fear	12	13	13	**62**	0.0001

Table 3 Emotional voice association rates on a motherese-trained model. PRA=Praise, COM=Comfort, PRO=Prohibition, ATT=Attention.

recorded by professional German actors. Further information on this voice analysis can be found in [49]. The emotional voice samples were tested against the MEI, and the GMM which output the highest probability was selected as the best match.

6.3 Results and discussion

6.3.1 Association of happy voices with flourishing, and sad voices with distress

The results illustrate the robot's empathy grounded in physical feelings. In Table 2, we can see that happy voices from the German database were associated with the physical flourishing state 90% of the time. Sad voices were associated with the distress state 84% of the time. This result is reasonable when noticing the similarities between the Praise (flourishing) and Happiness dynamics, versus Comfort (distress) and Sadness dynamics in Figure 8.

This suggests that a robot could develop a physically grounded association in response to happy and sad voices, by exposing the robot to comforting and praise motherese when it is in low-battery or high-battery states, respectively. This is interesting because this limited exposure mimics the sequence of motherese directed at very young infants: at 3 months, infants prefer and receive more comfort vocalizations. Then, at 6 months, they prefer and receive approval (praise) vocalizations [41].

This is a significant result, because it shows that exposure to motherese can result in adult-like empathy. Consider that early robot emotion systems, such as Kismet [11], were trained with motherese and were limited to recognizing motherese (specifically, motherese spoken by females). Similarly, emotional voice systems are trained with adult emotional speech to recognize adult emotional speech (e.g., [5]). Cross-dataset recognition is rarely attempted [49]. Our system is able to make the leap because of its use of high-level abstract features, recognizing male and female adult-directed emotional voices with training only with motherese.

Finally, it is important to note that these associations are not simply recognition systems. They are bidirectional, because the robot develops its emotional expression based on the input it receives. In contrast, the typical way to design a robot emotion system is to assemble many recognizers and hand-designed expression systems, which do not interact [11]. This is limiting because the robot is not adaptive: even if it is exposed to hundreds of hours of emotional interaction, it would not change the way it expresses emotion.

6.3.2 Associations of happy, sad, angry and scared emotional voices

What happens when a robot has been exposed not only to two types of motherese vocalizations, but four? Table 3 shows the output of the robot's recognition system which was trained with motherese and analyses adult-directed emotional voice. Our hypothesis is that a) happy voices are associated most strongly with praise b) sad voices associated with comfort c) anger voices associated most strongly with prohibition. As a preliminary hypothesis, it is not clear that fear (a negative emotion) would correlate with attention voices, because "attention bids" were described as a positive, playful interaction for caregivers and infants in [28].

In Table 3, we first notice that recognition rates of happiness and sadness drop, but that sadness is still associated with comfort at 65%, and happiness associated with praise at 54%, both at levels significantly higher than chance (Table 3). Happiness is sometimes associated with the prohibition condition, at 24%. Anger is not well associated, being confused with happiness. Surprisingly, fear voices are associated with attention motherese, to a high degree. It is interesting to note that infants prefer motherese vocalizations in a preset order: at 3 months, they prefer comfort vocalizations. At 6 months, they prefer approval (praise) vocalizations. Lastly, at 9 months, they prefer directive vocalizations [41]. The association rates of first comfort, then praise, and lastly prohibition parallel this infant progression.

In Figure 8, we can clearly notice the relationship between emotional voices and motherese. Here, the GMM means for both motherese and German emotional voice models are plotted for comparison purposes. In Table 4,

Motherese Emotional Voice	**PRA**	**COM**	**PRO**	**ATT**
Happiness	**.13**	.67	.20	.39
Sadness	.39	**.36**	.47	.39
Anger	.30	.72	**.28**	.37
Fear	.30	.39	.43	**.18**

Table 4 Euclidean distance between SIRE means of 1-mixture GMMs for motherese and emotional voice classes. PRA=Praise, COM=Comfort, PRO=Prohibition, ATT=Attention. Lower values indicate that the two classes are more similar. Distances in bold show the closest motherese profile for a given emotion in voice.

we show the same data in more detail, by calculating the Euclidean distance between the GMM means. We can see that happiness and praise reside very close to each other, as do fear and attention.

One way to interpret these results is that praise, comfort and attention motherese build corresponding elements of adult emotional intelligence. Notice that the praise, comfort and attention conditions have very clear correlates to happy, sad and fear voices (e.g., a fear voice means that the human is trying to attract their attention to something). Although we did not predict that the attention motherese condition would build any appropriate emotional reaction, the fear and attention SIRE profiles appeared to be very similar (Fig. 8 and Table 4). This could make sense, since during fear vocalizations by a caregiver, it would be necessary to attract the infants' attention of the danger. This being said, further human studies to explore this idea would be needed.

7 Limitations

We should note that the system is limited in the following ways. First of all, we did not address facial expressions, which are a major component of emotional expression. Since motherese is also accompanied by exaggerated facial expressions, however, it could likely be added to future experiments using the framework as described.

Secondly, we addressed only affective empathy. Cognitive empathy may be a complementary empathetic skill requiring higher levels of cognition, such as described in Kozima's related work in empathetic developmental robots [44]. Asada et al. also explore this idea of robot empathy based on a " knowledge base of experiences" and "logical understanding" [4].

Further, we did not address other aspects of emotion as defined in the introduction, such as action tendency and regulation. Practically, a robot should perform these higher level aspects; for example, a robot should not simply stay sad if we are sad, but rather regulate its emotions to entrain to a positive state together.

Finally, we can consider an even deeper modification of the robot's internal state during emotional contagion: if the robot has low battery, and a user expresses happiness to boost the robot's emotional state, might the robot believe it has higher battery than it actually does? Would energy resources be allocated differently? This can lead to interesting discussions of robot self-awareness and information hiding in an intelligent robot.

8 Conclusions and outlook

What are feelings? Why do we feel pain when watching others in pain? Ideas addressing the perplexing phenomenon of emotion have been suggested in scientific studies, but as roboticists building the system in detail, aiming for authentic experiences, we must demand or create better explanations. As stated by Richard Feynman: "What I cannot build, I cannot understand."

In this paper, we created a new and detailed explanation for affective empathy, based on evidence in human neuroscience and developmental psychology. The model proposed that feeling another's sadness as one's own physical pain is the emergent result of 1) a functional mirror system, insular cortex, and somatosensory cortex and 2) specific motherese interaction with empathetic human caregivers. Importantly, we created a working robotic implementation of a mirror system architecture as a proof of concept. Results showed that a statistically learning robot exposed to *praise* and *comfort* interactions develops physically grounded "gut feeling" associations in response to happy and sad voices. Further, the robot trained with infant-directed "attention bids" recognized adult fear voices.

By isolating the empathetic function in a robot, this paper raises important questions both about humans and robots. Is this how human empathy is developed, too? Could the lack of praise or comfort interactions lead to non-empathetic adults? We propose this model as a concrete framework towards better understanding of the human emotion system, for the advancement of both engineering and science.

References

1. Adolphs, R., Damasio, H., Tranel, D., Cooper, G., Damasio, A.R.: A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. The Journal of Neuroscience **20**(7), 2683–2690 (2000)

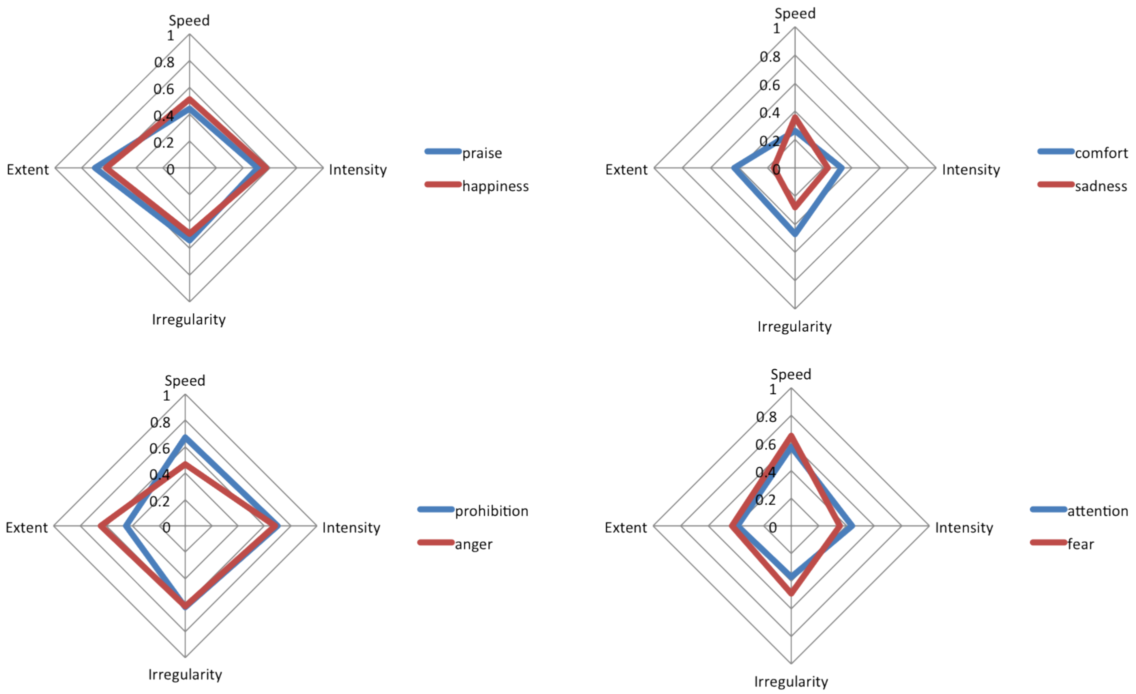

Fig. 8 Plotting the SIRE means of 1-mixture GMMs trained in each condition.

2. Adolphs, R., Tranel, D., Damasio, A.R.: Dissociable neural systems for recognizing emotions. Brain and Cognition **52**(1), 61–69 (2003)
3. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., Yoshida, C.: Cognitive Developmental Robotics: A Survey. IEEE Trans. Auton. Mental Dev., **1**1, 12–34.
4. Asada, M. Nagai, Y. Ishihara, H.: Why Not Artificial Sympathy?. In: International Conference on Social Robotics, 278–287. (2012)
5. B, S., Batliner, A.: Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. John Wiley & Sons (2013)
6. Baron-Cohen, S., Wheelwright, S.: The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. Journal of autism and developmental disorders, **34**(2), 163–175 (2004).
7. Baron-Cohen, S.: Zero degrees of empathy: A new theory of human cruelty. Penguin UK (2011)
8. Bernhardt, B.C., Singer, T.: The neural basis of empathy. Annual review of neuroscience **35**, 1–23 (2012)
9. Best, C.T.: Accommodation in mean f, during mother-infant and father-infant vocal interactions: a longitudinal case study. J. Child Lang **4**(7), 9–736 (1997)
10. Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L.: Imitation as a communication tool for online facial expression learning and recognition. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pp. 5323–5328. IEEE (2010)
11. Breazeal, C.L.: Designing Sociable Robots with CDROM. MIT press (2004)
12. Butti, C., Hof, P.R.: The insular cortex: a comparative perspective. Brain Structure and Function **214**(5-6), 477–493 (2010)
13. Calder, A.J., Keane, J., Manes, F., Antoun, N., Young, A.W.: Impaired recognition and experience of disgust following brain injury. Nature neuroscience **3**(11), 1077–1078 (2000)
14. Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., Lenzi, G.L.: Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. Proceedings of the National Academy of Sciences **100**(9), 5497–5502 (2003)
15. Clore, G.L., Palmer, J.: Affective guidance of intelligent agents: How emotion controls cognition. Cognitive Systems Research **10**(1), 21–30 (2009)
16. Craig, A.D.: How do you feel—now? the anterior insula and human awareness. Nature Reviews Neuroscience **10**, 59–70 (2009)
17. Cramer, H., Goddijn, J., Wielinga, B., Evers, V.: Effects of (in) accurate empathy and situational valence on attitudes towards robots. In: Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, pp. 141–142. IEEE Press (2010)
18. Damasio, A.: Descartes' error: Emotion, reason and the human mind. New York: Grossett/Putnam (1994)
19. Damasio, A.R.: Looking for Spinoza: Joy, sorrow and the feeling brain. Random House (2004)
20. Davis, M., Whalen, P.J.: The amygdala: vigilance and emotion. Molecular psychiatry **6**(1), 13–34 (2001)
21. De Waal, F.B.: Putting the altruism back into altruism: the evolution of empathy. Annu Rev Psychol **59**, 1–22 (2007)
22. Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G.: Understanding motor events: a neurophysiological study. Experimental Brain Research **91**(1), 176–180 (1992)
23. Dinstein, I.: Human cortex: reflections of mirror neurons. Current Biology **18**(20), R956–R959 (2008)
24. Edele, A., Dziobek, I., Keller, M.: Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. Learning and Individual Differences **24**, 96–102 (2013)

25. Fellous, J.M., Arbib, M.A.: Who needs emotions?: The brain meets the robot. Oxford University Press Oxford (2005)
26. Fellous, J.M., Ledoux, J.E.: Toward basic principles for emotional processing: What the fearful brain tells the robot. In: J.M. Fellous, M.A. Arbib (eds.) Who needs emotions?: The brain meets the robot. Oxford University Press Oxford (2005)
27. Fernald, A.: Intonation and communicative intent in mothers' speech to infants: Is the melody the message? Child development pp. 1497–1510 (1989)
28. Fernald, A.: Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. Child development **64**(3), 657–674 (1993)
29. Fernald, A., Simon, T.: Expanded intonation contours in mothers' speech to newborns. Developmental psychology **20**(1), 104 (1984)
30. Gleason, J.B.: Talking to children: Some notes on feedback. Talking to Children pp. 199–205 (1977)
31. Gonsior, B., Sosnowski, S., Buß, M., Wollherr, D., Kuhnlenz, K.: An emotional adaption approach to increase helpfulness towards a robot. In: Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pp. 2429–2436. IEEE (2012)
32. Grossmann, T., Striano, T., Friederici, A.D.: Crossmodal integration of emotional information from face and voice in the infant brain. Developmental Science **9**(3), 309–315 (2006)
33. Hatfield, E., Cacioppo, J.T.: Emotional Contagion. Cambridge University Press (1994)
34. Hein, G., Singer, T.: I feel how you feel but not always: the empathic brain and its modulation. Current Opinion in Neurobiology **18**(2), 153–158 (2008)
35. Iacoboni, M.: Imitation, Empathy, and Mirror Neurons. Annual Review of Psychology **60**(1), 653–670 (2009)
36. Iacoboni, M., Dapretto, M.: The mirror neuron system and the consequences of its dysfunction. Nature Reviews Neuroscience **7**(12), 942–951 (2006)
37. Ishihara, H., Yoshikawa, Y., Asada, M.: Realistic child robot Affetto for understanding the caregiver-child attachment relationship that guides the child development. Development and Learning (ICDL) **2**, 1–5 (2011)
38. Juslin, P.N., Simon, L., Daniel, V., Lars-Olov, L.: How does music evoke emotions? exploring the underlying mechanisms. In: P.N. Juslin, J.A. Sloboda (eds.) Handbook of music and emotion: Theory, research, applications. Oxford University Press (2010)
39. Juslin, P.N., Sloboda, J.A.: Handbook of music and emotion: Theory, research, applications. Oxford University Press (2010)
40. Kim, P., Feldman, R., Mayes, L.C., Eicher, V., Thompson, N., Leckman, J.F., Swain, J.E.: Breastfeeding, brain activation to own infant cry, and maternal sensitivity. Journal of Child Psychology and Psychiatry **52**(8), 907–915 (2011)
41. Kitamura, C., Lam, C.: Age-specific preferences for infant-directed affective intent. Infancy **14**(1), 77–100 (2009)
42. Koelsch, S., Fritz, T., von Cramon, D.Y., Müller, K., Friederici, A.D.: Investigating emotion with music: An fMRI study. Human Brain Mapping **27**(3), 239–250 (2006)
43. Kohler, E.: Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. Science **297**(5582), 846–848 (2002)
44. Kozima, H., Nakagawa, C., Yano, H.: Can a robot empathize with people? Artificial Life and Robotics **8**(1), 83–88 (2004)
45. Kwak, S.S., Kim, Y., Kim, E., Shin, C., Cho, K.: What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In: RO-MAN, 2013 IEEE, pp. 180–185. IEEE (2013)
46. Leite, I., Pereira, A., Castellano, G., Mascarenhas, S., Martinho, C., Paiva, A.: Modelling empathy in social robotic companions. In: Advances in User Modeling, pp. 135–147. Springer (2012)
47. Lim, A., Ogata, T., Okuno, H.G.: Converting emotional voice to motion for robot telepresence. In: Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on, pp. 472–479. IEEE (2011)
48. Lim, A., Ogata, T., Okuno, H.G.: Towards expressive musical robots: a cross-modal framework for emotional gesture, voice and music. EURASIP Journal on Audio, Speech, and Music Processing **2012**(1), 1–12 (2012)
49. Lim, A., Okuno, H.G.: The MEI Robot: Towards Using Motherese to Develop Multimodal Emotional Intelligence. IEEE Transactions on Autonomous Mental Development **(In press)**
50. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychol.: Development., Learn., Personal., Soc., **14**, 261–292 (1996)
51. Morrison, I., Bjornsdotter, M., Olausson, H.: Vicarious Responses to Social Touch in Posterior Insular Cortex Are Tuned to Pleasant Caressing Speeds. Journal of Neuroscience **31**(26), 9554–9562 (2011)
52. Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., Ball, T.: A Rapid Sound-Action Association Effect in Human Insular Cortex. PLoS ONE **2**(2), e259 (2007)
53. Nagai, Y., Rohlfing, K.J.:Computational Analysis of Motionese Toward Scaffolding Robot Action Learning. Autonomous Mental Development, IEEE Transactions on, **1**(1), 44–54 (2009)
54. Neumann, R., Strack, F.: "Mood contagion": The automatic transfer of mood between persons. Journal of personality and social psychology **79**(2), 211–223 (2000)
55. Nummenmaa, L., Hirvonen, J., Parkkola, R., Hietanen, J.K.: Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. NeuroImage **43**(3), 571–580 (2008)
56. Pereira, A., Leite, I., Mascarenhas, S., Martinho, C., Paiva, A.: Using empathy to improve human-robot relationships. In: Human-Robot Personal Relationships, pp. 130–138. Springer (2011)
57. Pohl, A., Anders, S., Schulte-Rüther, M., Mathiak, K., Kircher, T.: Positive Facial Affect – An fMRI Study on the Involvement of Insula and Amygdala. PLoS ONE **8**(8), e69,886 (2013)
58. Preston, S.D., De Waal, F.: Empathy: Its ultimate and proximate bases. Behavioral and brain sciences **25**(01), 1–20 (2002)
59. Reissland, N., Stephenson, T.: Turn-taking in early vocal interaction: a comparison of premature and term infants vocal interaction with their mothers. Child: care, health and development **25**(6), 447–456 (1999)
60. Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., Robinson, P.: Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. IEEE Affective Computing and Intelligent Interaction and Workshops, 1–6. (2009).
61. Rizzolatti, G., Craighero, L.: The mirror neuron system. Annual Review of Neuroscience **27**(1), 169–192 (2004)
62. Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.C., Cohen,

D.: Motherese in interaction: At the cross-road of emotion and cognition?(a systematic review). PloS one **8**(10), e78,103 (2013)
63. Shic, F., B. Scassellati: How Not to Evaluate a Developmental System. IJCNN, 52185225 (2006)
64. Schwarz, G.: Estimating the dimension of a model. The annals of statistics **6**(2), 461–464 (1978)
65. Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D.: Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain **132**(3), 617–627 (2009)
66. Sievers, B., Polansky, L., Casey, M., Wheatley, T.: Music and movement share a dynamic structure that supports universal expressions of emotion. Proceedings of the National Academy of Sciences **110**(1), 70–75 (2013)
67. Singer, T.: Empathy for Pain Involves the Affective but not Sensory Components of Pain. Science **303**(5661), 1157–1162 (2004)
68. Soken, N.H., Pick, A.D.: Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. Child development **63**(4), 787–795 (1992)
69. Trainor, L.J., Austin, C.M., Desjardins, R.N.: Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological science **11**(3), 188–195 (2000)
70. Turkle, S.: Alone together: Why we expect more from technology and less from each other. Basic Books (2012)
71. Velsquez, J. D., Modeling emotions and other motivations in synthetic agents, pp. 1015 (1997)
72. Walker-Andrews, A.S.: Infants' perception of expressive behaviors: differentiation of multimodal information. Psychological bulletin **121**(3), 437 (1997)
73. Watanabe, A., Ogino, M., Asada, M.: Mapping facial expression to internal states based on intuitive parenting. Journal of Robotics and Mechatronics **19**(3), 315 (2007)
74. Yamada, T., Hashimoto, H., Tosa, N., Pattern recognition of emotion with neural network, *in Proceedings of IECON*, pp. 183–187 (1995)

Dr. Angelica Lim received her B.Sc. in Computing Science from Simon Fraser University, and her M.Sc. and Ph.D. in Informatics from Kyoto University in 2014. She worked for Aldebaran Robotics Japan, and has interned at Google, Honda Research Institute Japan, and I3S-CNRS, France. Since 2006, she has been engaged in robotics and artificial intelligence, and is currently interested in signal processing, machine learning and developmental robotics for intelligent systems, particularly in the field of emotions. She was a Guest Editor for the International Journal of Synthetic Emotions, and has received various awards including CITEC Award for Excellence in Doctoral HRI Research (2014), NTF Award for Entertainment Robots and Systems at IROS 2010, and the Google Canada Anita Borg Scholarship (2008). She is a contributor to IEEE Spectrum Automaton Robotics Blog, and is a member of the IEEE.

Prof. Hiroshi G. Okuno received a B.A. and Ph.D from the University of Tokyo in 1972 and 1996, respectively. He worked for NTT, JST, Tokyo University of Science, and Kyoto University. He is currently a professor of Graduate School of Creative Science and Engineering, Waseda University amd a professor emeritus, Kyoto University. He is currently a professor of Graduate Program for Embodiment Informatics, Graduate School of Creative Science and Engineering, Waseda University, and a professor emeritus, Kyoto University. He was visiting scholar at Stanford University from 1986 to 1988. He is currently engaged in computational auditory scene analysis, music information processing and robot audition. He received various awards including the 1990 Best Paper Award of JSAI, the Best Paper Award of IEA/AIE-2001, 2005, 2010, and 2013, IEEE/RSJ IROS-2001 and 2006 Best Paper Nomination Finalist, and NTF Award for Entertainment Robots and Systems in 2010. He co-edited "Computational Auditory Scene Analysis" (Lawrence Erlbaum Associates, 1998), "Advanced Lisp Technology" (Taylor and Francis, 2002), and "New Trends in Applied Artificial Intelligence (IEA/AIE)" (Springer, 2007). He is a fellow of the Japanese Society for Artificial Intelligence, and a member of AAAI, ACM, ASA, RSJ, IPSJ, JSSST and JCSST.