Developing Robot Emotions through Interaction with Caregivers

Angelica Lim[1] and Hiroshi G. Okuno[2]
Kyoto University, Japan[1]
Kyoto University and Waseda University, Japan[2]

Abstract

In this chapter, we explore social constructivist theories of emotion, which suggest that emotional behaviors are developed through experience, rather than innate. Our approach to artificial emotions follows this paradigm, stemming from a relatively young field called developmental or 'epigenetic' robotics. We describe the design and implementation of a robot called MEI (multimodal emotional intelligence) with an emotion development system. MEI synchronizes to humans through voice and movement dynamics, based on mirror mechanism-like entrainment. Via typical caregiver interactions, MEI associates these dynamics with its physical feeling, e.g. distress (low battery or excessive motor heat) or flourishing (homeostasis). Our experimental results show that emotion clusters developed through robot-directed motherese ("baby talk") are similar to adult happiness and sadness, giving evidence to constructivist theories.

1 Introduction

> "Some of the most revolutionary ideas in brain science are coming from cribs and nurseries."
>
> – Patricia Kuhl

Are emotions innate? Recently, the popular Darwinian theory that basic emotions – such as happiness, sadness and anger – are "hard-wired" through evolution has been called into question. Psychologists have collected growing evidence (see reviews, (Mason & Capitanio, 2012; Camras & Shuster, 2013; Barrett, 2006)) that emotions may not in fact be completely a product of innate biology. Instead, "social constructivist" theories (Averill, 1980) point to experiences and environment as a prime factor for the development of emotions:

> "While there is little doubt that what we call fear, anger, and sadness refer to real (i.e., observable) phenomena and important parts of human experience, the weight of scientific opinion appears to be shifting away from the view that a few specific emotions are natural and universal kinds, laid down in the biology of humans and other animals (nature), in favor of a larger place for experience (nurture) in all emotions"– (Mason & Capitanio, 2012, p. 239)

Let us briefly illustrate this view with evidence from infant psychology, animal behavior and cultural emotion psychology.

In infant developmental psychology, (Camras L. A., 2011) has pointed out several phenomena that support the constructivist view. Firstly, emotional facial expressions were observed in infants where the emotion was not expected to occur: 5-7 month olds showed prototypical surprise expressions while bringing familiar objects into their months (Camras, Lambrect, & Michel, 1996). Secondly, emotional expressions were not observed in contexts during which the emotion should have occurred: 10-12 month olds in the visual cliff procedure rarely produced the fear expression, even though their other behaviors showed that they did in fact experience fear (Hiatt, Campos, & Emde, 1979). Finally, Camras and her colleagues also found that negative emotion classes (such as anger, sadness and fear) did not seem to differentiate well in infants as old as 11 months, suggesting that they all corresponded to a general negative "distress" affect. (Camras L. A., 2011).

Studies on atypical caregiver conditions in young animals also support the social constructivist view. In a study with rats, it was shown that when mother rats' maternal style contained more licking and grooming, their pups grew up to be less fearful, with decreased hormonal reactions to stressors (Kaffman & Meaney, 2007). In studies with rhesus monkeys, maternal separation early in life affected gene expression in brain regions controlling socio-emotional behaviors, with a correlation on the timing of the separation (Sabatini, Ebert, Lewis, Levitt, Cameron, & Mirnics, 2007). For example, monkeys separated from their mother at 1 week of age showed less expression of the gene GUCY1A3 (associated with social-seeking comfort behaviors), compared to the 1 month old separation condition.

Human studies on atypical early caregiving conditions also exist, though are more rare due to ethical issues. For instance, observations on postinstitutionalized (PI) children, such as those adopted from Eastern European orphanages after World War II, provide evidence that nurture is important for emotional intelligence. According to (Fries & Pollak, 2004), PI individuals had difficulty matching appropriate faces to happy, sad and fearful scenarios, yet were able to match angry faces just as well as controls. We refer the reader to further observations on the effect of early adverse rearing in work by Tottenham et al., e.g., (2011). In addition, temperament, while thought to have an inherent basis from birth, is not stable over the lifetime. In a short-term longitudinal study, (Calkins, 2002) found that infants who experienced negative parenting continued to show high anger/frustration levels, though it was not the case for infants who experienced positive parenting. Studies such as these emphasize the importance of emotional input early in life.

Psychologists studying emotions across cultures also observe variations that question the idea of universal emotion definitions. According to research by Tsai, ideal affect tends to differ between Eastern and Western cultures (Tsai, 2007). Individuals in Western cultures report “feeling good” as high-arousal positive (HAP) affect, whereas Eastern cultures prefer low-arousal positive (LAP) affect, even after controlling for self-reports of temperament and other individual differences (see (Tsai, 2007) for a review of experiments). As a simple illustration, (Hess, Beaupré, & Cheung, 2002) reported that a large proportion of Asian Canadians preferred smiles from 20-60% intensity, whereas European Canadians significantly preferred smiles from 80-100% intensity. In cross-cultural emotion behavior recognition experiments (Elfenbein & Ambady, 2003) noticed a 9.3% drop in accuracy when attempting cross-cultural facial judgments. A similar study on vocal cues observed a 7% drop (Juslin & Laukka, 2003). While cultural display “rules” have been suggested by Ekman (Ekman & Friesen, 1969) to account for these differences, details on how these are developed (and how display rules account for recognition and preference, in addition to expression) remains an open area of research.

Most interestingly for the topic of this book, this emerging “constructivist” perspective on emotion theory opens up a new avenue for artificial emotion systems. This is because emotion theories (Cornelius, 2000) such as Darwin’s Evolutionary theory, James’ Bodily Theory, and Cognitive Appraisal Theory have presented several practical challenges for roboticists. In the Darwin paradigm (Darwin, 1872/1965), we face the problem of ecological validity of the “basic” emotions (Ekman, 1992) – full-blown emotions often studied in psychology rarely appear in typical human-robot interaction scenarios. In the Jamesian view (James, 1884), without biological bodies, nervous systems, hearts, and so on, it appears impossible for robots to ever have emotions and feelings. Critics suggest that copying surface behavior of emotions, such as facial expressions and poses, do not “count” as real emotions, and their use in companion robots has been called unethical (Turkle, 2012). Appraisal Theory (Frijda, 1986) has been the most advantageous for creating emotional behavior in artificial agents (e.g. OCC (Ortony, 1990)), where large and complex rule sets define emotional states and expressions. Yet these hand-designed rules are difficult to design because the engineer must completely describe all possible scenarios in which emotional reactions might take place. Furthermore, a new rule set must be adjusted for each culture (e.g., in Japan, anger emotions should not be displayed in social contexts (Kitayama, Markus, & Kurokawa, 2000)).

Figure 1 Main concept behind statistical emotion development. The robot takes the place of the infant, and emotional associations are form as a result of a caregiver's interactions in situ. (Mother and Daughter by Ian Grove-Stephensen https://flic.kr/p/56bERd under CC BY 2.0)

A developmental paradigm suggests that a learning entity could develop emotions on its own – emotional expression, recognition, triggering events, and so on – if exposed to the right environment. This is related to the concept of epigenetics, which describes mechanisms by which the environment can program "enduring effects on gene expression and cellular function" (Meaney & Ferguson-Smith, 2010). The relatively new field of "epigenetic robotics", also called "developmental robotics", has been developed in the last decade under this framework (Asada, et al., 2009). It is likely that the latest social, humanoid forms of robots such as Affetto (Ishihara, Yoshikawa, & Asada, 2011) and NAO[i] were key to this revolution.

To date, only two previous studies have used the developmental paradigm to ground artificial agent's emotional expressions. In the first study of its kind, Watanabe et al. proposed the "intuitive paradigm", in which the parents mimicked a virtual infant's facial expression, to associate the facial expressions with the robot's internal state (Watanabe, Ogino, & Asada, 2007). The authors used the concept of Hebbian learning, to create associative links for later emotion recognition. Boucenna's study followed a similar strategy, using a physical robot (Boucenna, Gaussier, Andry, & Hafemeister, 2010). These are landmark works using the caregiver paradigm to link external emotional stimuli to expression, yet some major conceptual challenges still remain. In particular, since the era of Breazeal's emotional robot Kismet (Breazeal, 2004), the definition of robot "feeling" has never been tackled.

In this chapter we discuss a new emotion system called MEI, as an example study for what could be created under the social constructivist theory of emotion in the developmental robotics field. We will first suggest some pre-requisites: the statistical learning system architecture (brain), the physical condition of the robot (gut feeling) and learning process (environment). Importantly, we will define the concept of robot feeling. We then will describe the system in action – learning through interaction with caregivers in a naturalistic scenario. By inspecting the resulting models ("looking into the brain") and performing vocal recognition and expression tests, we suggest that MEI has achieved differentiated emotion representations and developed a basic form of emotional intelligence called "core affect".

2 General Overview

How do infant emotions develop into adult emotions? We propose that motherese interactions like the one in Figure 1 may serve as the basis of acquisition of emotion. Motherese is “baby talk” between a caregiver and infant at close proximity (Fernald, 1989). This exaggerated speech typically co-occurs with exaggerated facial displays (Soken & Pick, 1992), and is known to exist in all cultures of the world (Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies, & Fukui, 1989). It has been established that the highly exaggerated form of speech is used to aid the child in the acquisition of language (Kuhl, 2004). Furthermore, studies comparing adult-directed emotional speech with motherese show that motherese is also highly correlated with emotional speech, with robust differences across the emotions (Trainor, Austin, & Desjardins, 2000). As such, a recent review has suggested that motherese may exist for the development of emotions (Saint-Georges, et al., 2013). The universal social phenomenon of motherese therefore serves as the basis for our emotion development work.

2.1 Issues

The goal in developmental robotics is to make a robot that learns just as a child does. Yet this is an enormous task, to say the least. To encourage others in the developmental robotics or social constructivist paradigm to better formalize their individual contributions, we propose the following method of formalizing their assumptions and contribution.

In a developmental robotics study, one should specify which elements are assumed to be “innate”, and which are learned through environment. One way to specify this is by grouping them into Before (innate and previously learned), During Learning (environment) and After (the learned result). Another way to specify starting capacities is to approximate the robot’s “starting age” – to what human period do we assume the robot has already “grown”, and what milestones has it already reached? After the interaction, what new milestone or developmental age has it achieved? Finally, it is useful to define the human functional equivalents for each component of the system architecture. We illustrate this formalization method with such an overview for our system.

3 Overview of solution

We provide the following outline to clarify the goals of the present study, in which the robot develops *core affect*: “Core affect is a neurophysiological state that underlies simply feeling good or bad, drowsy or energised. Psychological construction is not one process but an umbrella term for the various processes that produce:

a. a particular emotional episode’s ‘‘components’’ (such as facial movement, vocal tone, peripheral nervous system change, appraisal, attribution, behaviour, subjective experience, and emotion regulation);
b. associations among the components; and
c. the categorisation of the pattern of components as a specific emotion.” (Russell, 2009, p. 1259)

In particular, we focus on (a) the components of *vocal tone*, *movement* and *subjective experience/feeling* (b) associations among these components and (c) categorization of the pattern of components as a specific emotion. In the present study, we look especially at the first emotions observed in the infant: happiness and sadness (Sigelman & Rider, 2010).

3.1 Before Interaction – System Requirements (Neonate)

We begin with several basic “innate” requirements for our robot system, those with functional parallels in a newborn infant.

	Requirements	Human equivalent	Robot equivalent	Age developed
1	A statistical learning neural system.	Brain	CPU, storage, machine learning program. More details on the learning system in Sec. 4.1.	At birth
2	A somatosensory system	Sensory system for internal functioning of gut, viscera, temperature, etc.	Battery level sensor and temperature sensor (Sec. 4.2).	At birth
3	A distress signal when the somatosensory system indicates a problem.	Crying to signal hunger, cold, etc.	Auditory distress signal (Sec. 4.5).	At birth

Next, we list the requirements that might be learned simultaneously or prior to the focus of this study, for instance within the neonatal stage (less than 4 weeks old).

	Requirements	Human equivalent	Robot equivalent	Age developed
4	An extrasensory processing system	Ears, eyes, skin and related nervous systems	Microphones and video camera (touch sensors, in the future) (Sec. 4.3).	~6 months
5	A mirroring system	Mirror neurons and pre-motor system	Mirror module (Sec. 4.4).	~4 weeks?

We defined two separate sections of "innate assumptions" and "previously learned requirements" because (4) and (5) are available, but not fully formed at birth. For example, although fetuses have auditory abilities even before birth (Lecanuet & Schaal, 1996), some aspects of hearing such as frequency and temporal resolution only reach adult levels at 6 months (Werner, 2002), and visual acuity reaches adult levels of 20/20 at 6 months of age (Sokol, 1978). Additionally, we do not yet know to what extent the human mirror system is formed and developed at birth (Craig, 2009). It has only been established that neonates (< 4 weeks) imitate tongue protrusions (Nagy, Pilling, Orvos, & Molnar, 2013). For this reason, we seek parsimony in the abilities placed in the "innate" section.

Let us illustrate these requirements with an example. Under these assumptions, we can imagine then that, for example, a smartphone with a battery level warning beep already has (2) and (3), possibly (4). With the right software, it could have (1). Yet, because of its lack of human-like embodiment and motor system, the mirroring system in (5) would be a difficult requirement to fulfill.

3.2 During Training – Environment (Neonate to 6 months old)

Now, we define the environmental input to which we will expose system with the above requirements: motherese. Motherese can be classified into categories, where the vocal tones of the caregiver's speech depend on the communicative intent to the infant (Fernald, 1989), including:

- **Comfort**: slow, falling pitch contours

- **Approval / Praise**: exaggerated rise-fall (bell-shaped) pitch contours
- **Attention bid**: quick, rising contours
- **Prohibition**: low pitch, high intensity and short

Infants between birth and 3-months old prefer *comfort* vocalizations most, according to (Kitamura & Lam, 2009), and it is the kind of motherese that caregivers produce most at that age. At 6 months, caregivers produce more *approval/praise* vocalizations, and infants' preference also shifts to approving tones. At 9 months, the preferences shift to *directive* tones.

One may wonder if an interactive robot is necessary to elicit motherese vocalizations. Wouldn't laboratory recordings and offline training be sufficient? In fact, it is not so easy – it has been shown that mothers are unable to reliably produce motherese in front of a microphone (Fernald & Simon, 1984). An infant – or a machine with the look and behavior of an infant – is important in eliciting the necessary environmental input (Ishihara, Yoshikawa, & Asada, 2011).

We propose that these motherese interactions create affective associations as follows. A typical caregiver reacts to a distress signal (crying) through comfort motherese. Between birth and 3 months of age, the co-occuring comfort tones and physical distress create an association between sadness-like sounds and a negative physical state. Between 3 and 6 months, a caregiver displays approval and praise when the infant is in good health, creating a positive bodily association with happiness-like sounds. We posit that the other types of motherese (prohibition, attention) teach infants further emotional associations in context, for example, prohibition tones associated with the context of being stopped from achieving a goal.

3.3 After (~6-8 months)

The final system, after caregiver interaction, should show an increase in *core affect* emotional capabilities:

- **Components**: emotional voice, movement, feeling of happiness and sadness
- **Association between components:**
    - *Feeling* to *expression*. The robot's subjective feeling of happiness should engender vocal tone and movement similar to happiness, and sad feelings should generate sad vocal tones and movement.
    - *Expression* to *feeling*. The robot should associate happy voices with a positive physical feeling, and sad voices with a negative physical feeling.
- **Categorization**: Differentiated happiness and sadness in the neural model.

Now that we have specified the Before, During Training and After phases, we describe the underlying system architecture, designed based on functional equivalents in the human brain.

3.4 Neural Architecture

In this section, we give details on the neural architecture we build into the robot. The three modules we identify, with associated human neural equivalents are as follows, as shown in Figure 2:

- **A mirror system:** represents the action of another human, and can induce eventual motor imitation (Premotor cortex) (Iacoboni & Dapretto, 2006)
- **A gut feeling module**: the module receiving signals of bodily pleasure or pain, e.g. battery level or motor heat (Somatosensory cortex) (Damasio A. , 1994)
- **An associative module:** associates the outputs of the above – action representation and a corresponding bodily feeling of pleasure or pain (Insula) (Craig, 2009)

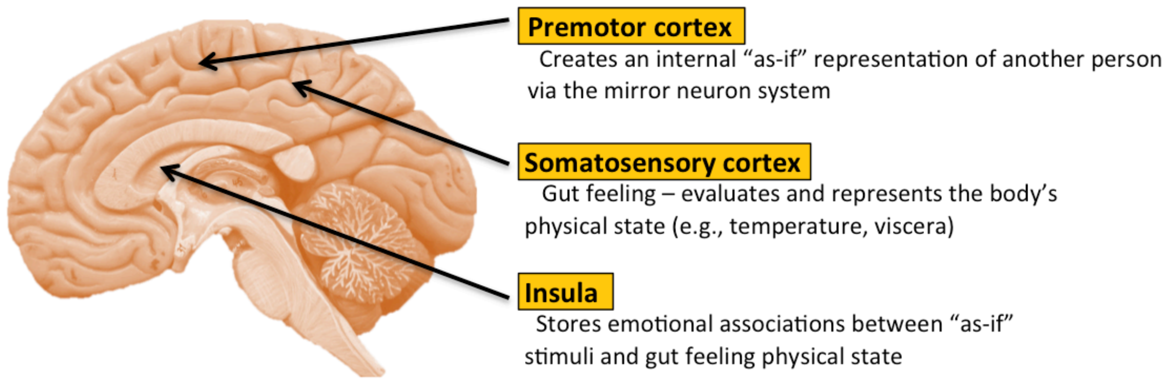

Figure 2: Proposed developmental emotion architecture. The basis of an artificial emotion system is in brain areas related to a) embodied feelings (somatosensory cortex) b) mirror representations of others (premotor cortex), and an associative lookup creating a link between them (insula). (Derived from Brain https://flic.kr/p/9UwYi by GreenFlame09 under CC BY 2.0)

Let's briefly examine the three brain areas that are key to defining emotion in our system:

3.4.1 The pre-motor cortex

Mirror neurons in the premotor-cortex been proposed as a critical step towards simulating and understanding of the mental states (including emotions) of others. In essence, we map actions of an observed person to internal representations of ourselves doing the same action. Simple visual observation of an action incites premotor activity in the brain (Rizzolatti & Craighero, 2004). For example, watching another grasp an object activates one's own premotor areas for grasping. Auditory input also achieves neural mirroring: neurons in the monkey premotor cortex discharge both when it performs an action and when it hears the related sound (Kohler, 2002).

Damasio called this internal representation 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." (Damasio A. R., 2004).

Therefore, our emotional robot system should include an artificial pre-motor cortex, containing an internal representation of the other person's body state.

3.4.2 The insula

The insula has been suggested to lie at the heart of the association between action representation and emotion (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). The insula is a region deep in the brain that reads one's body condition by way of visceral and interoceptive sensors (e.g. heat, cold, pain, muscle ache sensors) that send information to the insula. It has been associated with many behaviors such as drug cravings, feeling pain, maternal love, listening to music and empathizing with others (Craig, 2009). It is where a bad smell is transformed into disgust (Calder, Keane, Manes, Antoun, & Young, 2000) or a caress into pleasure (Morrison, Mjornsdotter, & Olausson, 2011), and is active when a mother hears her baby cry (Kim, et al., 2011), or when looking at a happy face (Pohl, Anders, Schulte-Ruther, Mathiak, & Kircher, 2013). It is active when empathizing for others' pain as well as when actually feeling pain (Singer, 2004), suggesting that the affective component of pain "feeling pain" is decoupled from the sensory component–the pain itself.

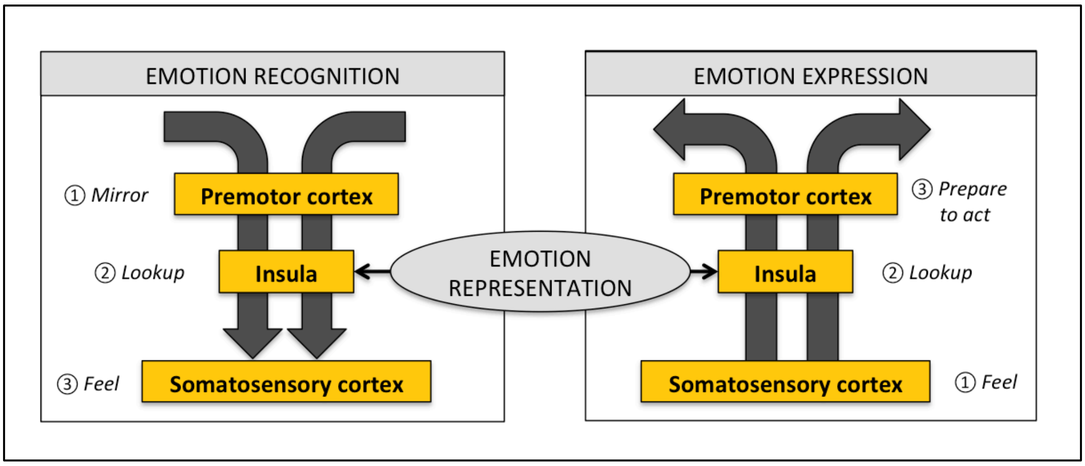

Figure 3: Proposed developmental emotion architecture (cont'd). The same 3-module architecture, containing an emotion representation in the insula, is responsible for both emotion recognition and emotion expression.

To summarize, Damasio and others have suggested that the insula maps visceral states that are associated with emotional experience (Damasio A. R., 2004; Craig, 2009; Singer, 2004). Based on this evidence, it seems an artificial insula is an integral part of a robot emotion system, to associate physical "gut" feelings (Sec. 3.4.3) with emotional body representations (Sec. 3.4.1).

3.4.3 The somatosensory cortex

In this section, we attempt to define feelings. Damasio's Somatic Marker Hypothesis suggests that feelings are an association of stimuli to visceral (and musculoskeletal) pleasure or pain: "Feelings are […] first and foremost about the body, that they offer us the cognition of our visceral and musculoskeletal state. […] Body images give […] a quality of goodness or badness, of pleasure or pain." (Damasio A. , 1994, p. 159)

In the brain, the somatosensory cortices (from the Greek root *soma*, meaning body) are responsible for sensing this internal state of pleasure or pain. They sense the body's internal state including viscera (e.g. internal organs like the heart, stomach, or lungs) and joint position, as well as external senses of touch, temperature and pain. Indeed, when the somatosensory areas are damaged, patients do not show normal signs of despair or panic (Damasio A. , 1994)

In short, the somatosensory cortex contains a "gut feeling" representation of the body's internal state, either flourishing or in distress. We consider that this gut feeling is linked to the emotional body representations (Sec. 3.4.1) within the insula (Sec. 3.4.3). And, to quote Damasio, "the critical, formative set of stimuli to somatic pairings is, no doubt, acquired in childhood and adolescents." (Damasio A. , 1994, p. 179)

In Figure 3, we illustrate the tight integration of the above three components. Unlike many robotic emotion systems with specialized systems for recognition and expression (e.g. Kismet (Breazeal, 2004)), the same system can perform both emotion recognition and expression, as illustrated by the two columns in Figure 3. Specifically, it can perform emotion recognition by first mirroring the person, then looking up the closest state and associated feeling. The latest neuroscientific models support this: according to "shared-substrates" models of emotion recognition, we understand other's emotions by first making an internal simulation of the other (Heberlein & Atkinson, 2009). Secondly, the architecture can perform emotion expression by first feeling a certain physical state, then expressing it by looking up the associated expression and preparing to act. Interestingly, the premotor cortex is only the preparation to act; possibly, the act of expressing the final emotion could be further mediated by the pre-frontal cortex and cognitive controls; for example, suppressing a smile in a socially inappropriate situation.

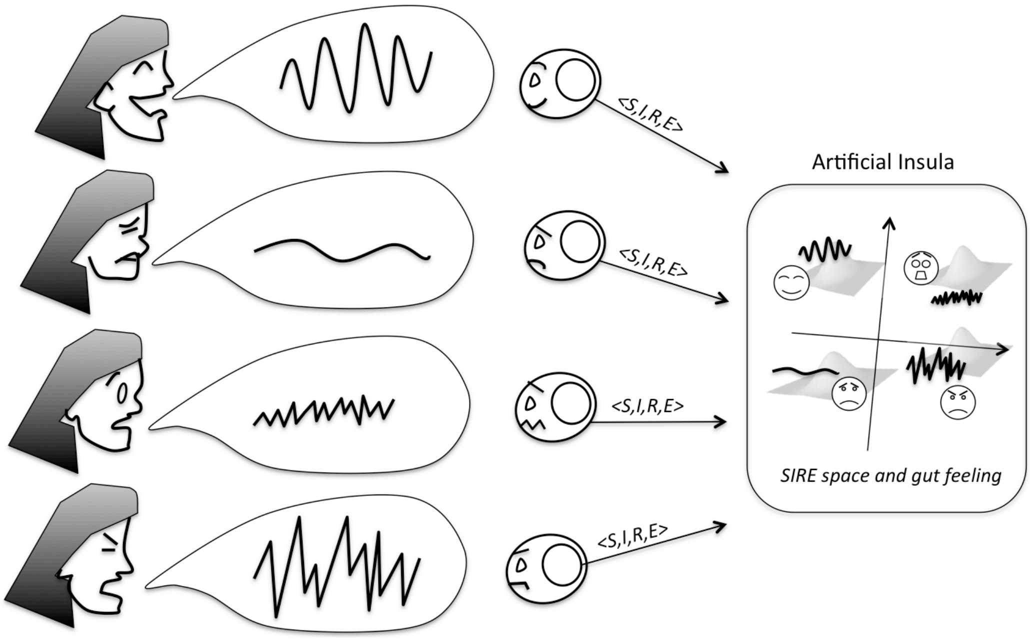

Figure 4: Illustration of the MEI system receiving affective input and clustering in the artificial insula to create differentiated classes.

4 MEI System Implementation

Based on the evidence presented above, we implemented the following five modules in a robot we call MEI – Multimodal Emotional Intelligence, as described in (Lim & Okuno, in press) and (Lim & Okuno, in prep.). The system was implemented in the Python programming language on the NAO robot, using HARK[ii] real-time audio processing technology, and machine learning algorithms as described below. We summarize the system here.

4.1 Statistical learning neural system (Artificial Insula)

The purpose of this module is to create associations between the robot's "gut feelings" (Sec. 3.4.3) and internal motor representations (Sec. 3.4.1). We use a Gaussian Mixture Model (GMM) (Bishop & Nasrabadi, 2006), which is a statistical learning model typically used for clustering and recognition, as illustrated in Figure 4. Gaussian Mixture Models are typically trained with large amounts of data, and the resulting model represents the data, with peaks around commonly observed data, and valleys for uncommon occurrences.

We chose the GMM to represent emotion clusters for many reasons. Firstly, GMMs are widely used in the field of audio and vision recognition. Essentially, a novel input is evaluated in the GMM, and the cluster with the highest score provides the recognition result. Secondly, unlike other recognition techniques like Support Vector Machines (Burges, 1998) or K-means (Kanungo, Mount, Netanyahu, Piatko, Silverman, & Wu, 2002), it can be used for expression. This is because the model represents a probability space that can be sampled – data around the "peaks" are more likely to be picked for expression, compared to the data in the valleys. Additionally, the samples will be non-repetitious, which conveniently represents the "noisy" nature of human expression. Finally, a GMM can be used for representation. We can inspect the peaks, known as the GMM means, which represent a "prototype" for that cluster.

Let us illustrate the method with an example. If the robot hears a comforting voice that is 1.0 syllables per second, then another comforting utterance at 1.4 syllables per second, it creates a Gaussian cluster that approximates a "prototypical comfort voice" with a mean speed of 1.2 syllables/sec and standard deviation of 0.2. Consider that the GMM space can represent dimensions other than speed, such as pitch range, etc., so the "prototype" can be multi-dimensional. This allows the mean to encapsulate not only speed, but also values for pitch, say 40 Hz on average. These other dimensions will be discussed in Sec. 4.3.

In addition, the system automatically splits clusters when the variance (related to standard deviation) of a cluster becomes large. We do this by calculating the Bayesian Information Criterion (BIC) (Schwarz, 1978) and finding the optimal number of clusters to represent the data (Lim & Okuno, in press). For instance, if a sad voice cluster begins to include both *high intensity* grief voices and *low intensity* comfort voices, the system may create 2 clusters instead of one. In this way, the learning module develops and differentiates as it receives more input.

Interestingly, the functions proposed by our GMM method sounds similar to the Differentiation and Dynamical Integration (DDI) of perspective of emotional development proposed by developmental psychologist Camras (2011). Our automatic creation of emotional clusters is similar to her concept of "attractor states". Initially, emotions are distinguished into positive or negative attractor states, but then further differentiation results in new attractor states corresponding to what we often call discrete emotions (e.g., negative affect splits into anger, sadness and fear).

One practical advantage of the GMM over the DDI model may be that the GMM is highly implementable. In experiments, we use the implementation from Scikit-learn machine learning library (Pedregosa, et al., 2011) written in the Python programming language, along with our own modifications for splitting clusters based on BIC scores. Details are available in (Lim & Okuno, in press).

4.2 Somatosensory system (artificial somatosensory cortex)

The purpose of this module is to represent the robot's internal "gut" feelings. Feelings are an integral part of an emotion system, yet to our knowledge, no other artificial emotion system has attempted to include it. For instance, the difficulty may stem from the preconception that feelings are an abstract concept. In fact, our definition of feelings is intrinsically linked to the fact that a robot has a body in the physical world.

We define the robot's feeling based on Damasio's idea of "flourishing" and "distress" (Damasio A. , 1994). This simple module checks for physical homeostasis in the robot:

- Gut feeling is set to *flourishing* if the robot's body is in homeostasis, i.e., the temperature of the joint motors is not too hot and not too cold, and if the battery percentage is over a certain value
- Otherwise, the gut feeling is set to *distress* if the robot is out of homeostasis, e.g., an arm motor is too hot

In experiments, the above was implemented using the NAO robot's Naoqi interface to the battery and arm temperature sensors.

4.3 Extrasensory processing system

The purpose of this module is to process and simplify the environmental information. In our current work, we process incoming information according to the SIRE model, as explained in (Lim, Ogata, & Okuno, 2012). In short, the SIRE model is a way to reduce complex signals such as voice, movement and music into a simple set of features. SIRE stands for Speed, Intensity, Regularity and Extent. For example, when processing speech signals, we can extract an utterance's *speech rate*, say, 2.0 syllables per second, and normalize it to fast (1.0), slow or stopped (0.0), or somewhere in between (between 0.0 and 1.0). This is also known as the utterance's Speed. Similarly, the concept of Speed within an arm s also exists: *velocity*. In the SIRE paradigm, each of speed, intensity, regularity and extent is represented as a value between 0 and 1, as follows:

- **Speed**: slow (S=0.0) vs. fast (S=1.0)
- **Intensity**: gradual (I=0.0) vs. abrupt (I=1.0)

- **Regularity**: rough (R=0.0) vs. smooth (R=1.0)
- **Extent**: small (E=0.0) vs. large (E=1.0)

Explicit mappings are described in (Lim, Ogata, & Okuno, 2012) and (Lim & Okuno, in press).

The point of representing features in this simplified manner is to use as minimal a representation for emotion as possible. For instance, our work has shown that a voice is considered sad when it is slow, low intensity, regular, with small pitch range (extent), S=0.1, I=0.4, R=0.7, E=0.4 (Lim, Ogata, & Okuno, 2012). A voice is perceived as happy when it is fast, not too intense, slightly irregular and with a rather large pitch range, S=0.7, I=0.2, R=0.2, E=0.7. Furthermore, gestures that followed these patterns of SIRE dynamics were perceived in similar ways. Therefore, we need not process and store more information than SIRE to represent these two states, which is preferable to aid in visualization and understanding of our models. In this paper we focus on voice and arm movements, but in the future we should include other concise feature representations, such as facial configurations (eyebrow angle, mouth corner changes) or SIRE based on touch.

In experiments, the above concepts were implemented using a PlayStation Eye microphone as input, and the HARK[ii] real-time audio processing system to extract the SIRE values.

4.4 Mirroring system (artificial pre-motor cortex)

The mirroring module has two purposes:

2. To create the robot's internal "as-if" representation, a simulation of its caregiver
3. To generate motor output (voice, arm movements) that incites interaction from caregiver.

Firstly, for (1), the mirroring system creates an internal "as-if" representation of its environment. The representation is essentially four numbers, representing SIRE. For example, if the caregiver speaks with slow speech, it entrains the robot's internal representation to also be slow (e.g., S=0.1). As shown in Figure 3, the robot first creates an internal "as-if" representation in the pre-motor cortex, and this internal "as-if" SIRE representation is used to create emotional clusters (Figure 4).

Secondly, for (2), motor output is generated via the artificial pre-motor cortex. Just as SIRE can be used for analyzing input, it can be used for generating output. For example, if a robot's output SIRE state has S=0.8, means the robot will speak fast, with fast gestures. In (Lim & Okuno, in prep.), we defined the formula for entrainment, which is a gradual synchronization of the SIRE dynamics between the human's input and the robot's output. For example, when the caregiver speaks to the robot with slow speech, it entrains the robot's SIRE motor output to also be slow. The actual output depends also on the robot's current feeling and the robot's previous SIRE dynamics, however. For instance, a robot that is very low on battery will be more difficult to soothe. Details are given in (Lim & Okuno, in prep.).)

Although not yet formally tested, we consider the robot's motor feedback in (2) as an essential part of eliciting realistic motherese from the human. For example, if the caregiver speaks to the robot with a happy voice, and the robot returns a similarly happy voice, the caregiver should become more aroused. Thus, a positive feedback loop is created.

In experiments, we implemented the mirror system using Naoqi ALMemory variables, Python code to implement the entrainment formula, and mappings from SIRE to the NAO robot's Naoqi interface to motor controls, as described in (Lim, Ogata, & Okuno, 2011).

4.5 Distress signal

The purpose of this module is to incite interaction from caregiver. The module continuously checks the robot's current gut feeling, and sends a distress signal if the feeling is *distress* (low battery or hot motors). In our experiments, we set the distress signal to mimic an infant's cry at birth: a high intensity sound, with a large pitch range and a very regular timing. In practice, we set the robot's Text-To-Speech (TTS) system to repeat the syllables "*ma ma ma*" and set SIRE = 0.9 for all four parameters.

In experiments, we used the NAO robot's Acapela[iii] TTS system and markup to modify the dynamics of the robot's voice.

5 Evaluation of the system

Let us recall our formalization in Section 3.3, in which we defined some ways to evaluate the system after interaction with a human. After a caregiver interaction, we hypothesize that the robot system should have acquired these three aspects:

1. **Categorization**: Differentiated happiness and sadness in the neural model. Ability to recognize adult happiness and sadness in the voice.
2. **Expression to feeling.** The robot should associate happy voices with a positive physical feeling, and sad voices with a negative physical feeling.
3. **Feeling to expression**. The robot's subjective feeling of happiness should engender vocal tone and movement similar to happiness.

5.1 Training Conditions

We recruited 6 fluent English speakers from Western countries (3 female, with a mean age of 29.8 years old). The participants were introduced to the robot and we asked them to say the robot's name ("Mei Mei") in two different scenarios, mimicking scenarios that happen in the first 6 months of a human infant's life:

- **Comfort**: Mei Mei is crying (SIRE = 0.9). Soothe and comfort her by saying her name.
- **Praise**: Mei Mei is no longer crying (SIRE = 0.1). Praise her and make her feel loved by saying her name.

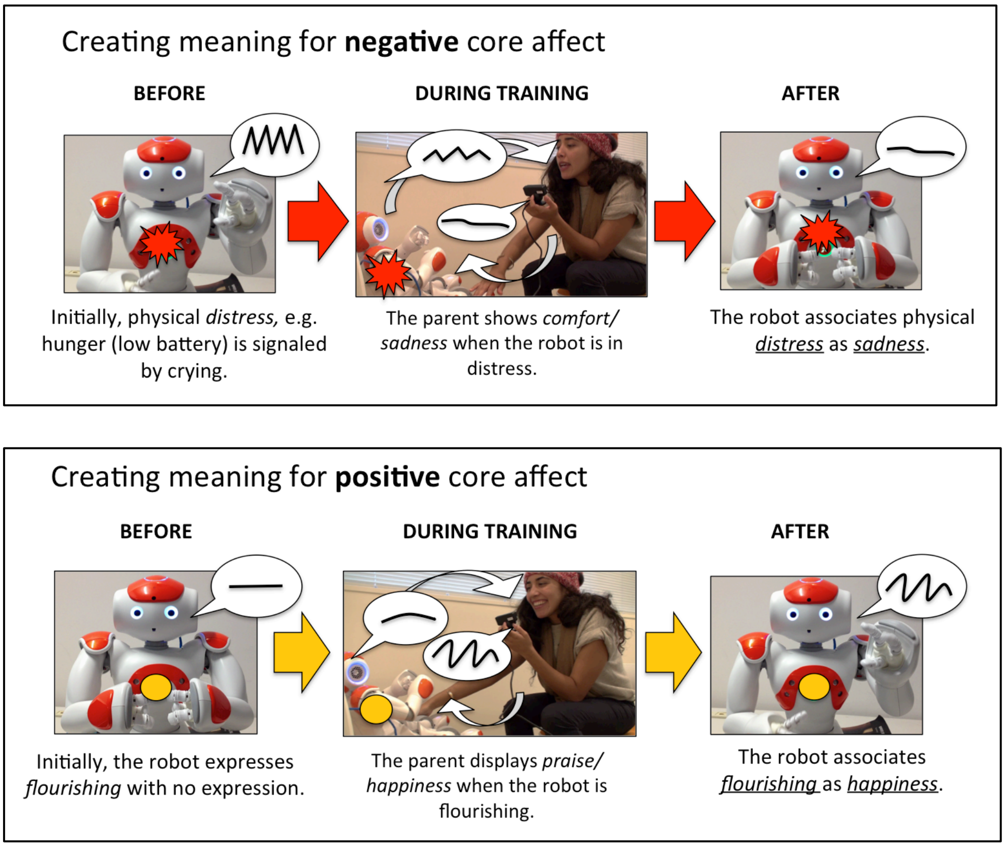

Figure 5: The creation of meaning for negative (above) and positive (below) core affect. Similar to how a newborn expresses distress such as hunger and cold by crying, the robot emits a high intensity signal when in physical distress. The caregiver regulates the expression using a comforting voice and face (~3 months in infants). After caregiver training, the robot has learned to associate the physical distress with the comfort/sadness dynamics. A similar association is made with positive physical flourishing state and praise/happiness dynamics.

The process is illustrated in Figure 5. During the comfort condition, the robot's feeling was set to distress, which engendered a vocal distress signal along with arm movements (SIRE = 0.9). During the praise condition, the robot moved its arms but did not vocalize, and the robot's feeling was set to flourishing (SIRE = 0.1 for all parameters). At all times, the robot continuously expressed its current SIRE state by gesturing with its arms. For example, when it was crying in distress (SIRE = 0.9 for all parameters), its gestures were initially fast, intense, regular, and large. When it was in a flourishing state, its gestures were initially slow, not intense, irregular and small. (SIRE = 0.1 for all parameters).

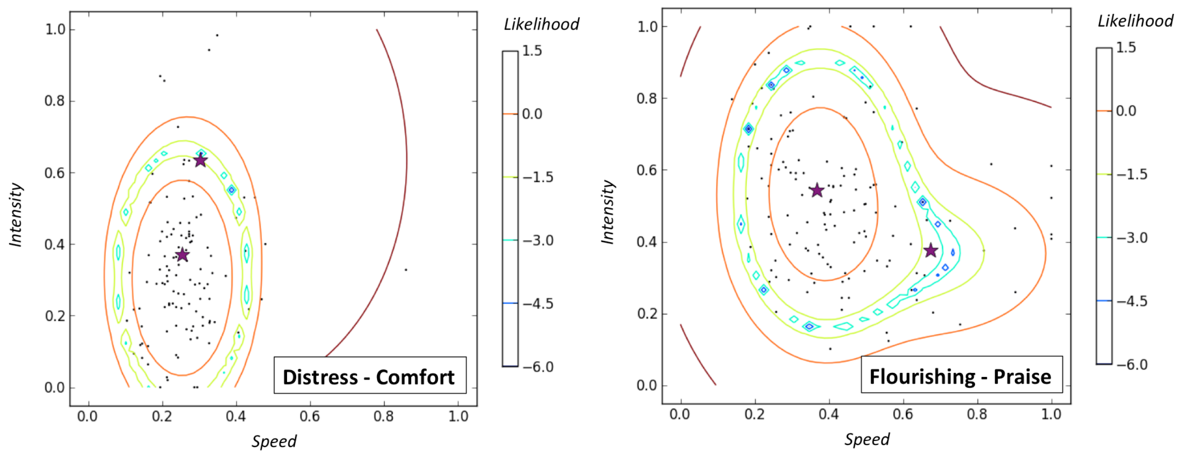

Figure 6: Trained emotion representations in artificial insula, associated with physical distress and physical flourishing. Visualization using only two of the four SIRE dimensions–speed and intensity. We can notice the clusters for comfort (cluster peaks denoted by ★) are generally slower than the clusters for praise.

5.2 Results and discussion

The interactions resulted in 128 praise utterances and 114 comfort utterances, which were used to train the robot's emotion model.

Categorization. The resulting model, partially shown in Figure 6, shows differentiation between conditions of distress and flourishing. We can already see differentiation in the two dimensions of speed and intensity, and further differentiation could be seen if it were possible to visualize all four SIRE dimensions. Another way to interpret the model is to plot the Gaussian means (the peaks denoted by a star, in Figure 6), where the means represent the 4-dimensional "prototypes". For example, the trained model produces prototypical *flourishing-praise* values as SIRE=[0.4, 0.5, 0.5, 0.7], and prototypical *distress-comfort* SIRE value as SIRE=[0.3, 0.3, 0.5, 0.4] (Lim & Okuno, in prep.). This intuitively makes sense, because praise voices have larger pitch ranges than comfort voices (praise E=0.7, comfort E=0.4). Further results showing Gaussian mean differentiation of attention and prohibition conditions, and their clear similarities to fear and anger voices, can be seen in (Lim & Okuno, in prep.).

Expression to feeling. As we report in (Lim & Okuno, in prep.), the model was able to associate 90% of happy vocal utterances with flourishing, and 84% of sad vocal utterances with distress. These results came from experiments testing the recognition ability of the motherese-trained GMM model, when exposed to 71 "happy" and 62 "sad" adult emotional voices from the German dataset Emo-DB (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005). In other words, happy voices induced higher likelihood scores in the "flourishing-praise" cluster than in the "distress-comfort" cluster. Similarly, sad voices resulted in higher likelihood scores in the "distress-comfort" clusters. In (Lim & Okuno, in prep.), we suggest that this is akin to affective empathy or "feeling another's pain": the robot performs internal mirroring (in this case of vocal dynamics) and makes a learned association with physical distress.

Feeling to expression. It may appear that we have only made a simple model of motherese and emotional voices. However, we insist that this model is also generative. In other words, the *motherese received by the robot affects its own expression*.

Our previous work has shown that our system can reliably express happiness and sadness through voice, gesture and gait, given emotional voice training alone (Lim & Okuno, in press). Evaluations were performed by 20 participants rating the robot using Mehrabian's Pleasure-Arousal-Dominance (PAD) scale (Mehrabian, 1995), following similar perceptual experiments in (Lim, Ogata, & Okuno, 2012). Perception experiments for the motherese-trained system described here are still underway, and we expect a similar result.

Intuitively, this reflects the idea that the caregiver's own emotional expressions are "transmitted" to their infant. Among humans, infants of depressed mothers continue to show depressed behaviors to other adults (Field, et al., 1988). Furthermore, early negative maternal parenting styles predicted greater increases in negative behaviors of the child later in life (Calkins, 2002).

Finally, our developmental robot system could explain Ekman and Friesen's "display rules" in a statistical manner. Our theory could explain why "feeling good" is associated with high arousal in Western cultures, and low arousal in Eastern Cultures: the parents simply entrained the flourishing states to different levels. Training the robot with an Eastern and Western caregiver and inspecting the cluster means could verify this statistical differentiation. Additionally, it is known that American motherese contains more extreme pitch modifications than British motherese (Shute & Wheldall, 1989). We could "raise" the robot with caregivers from two different cultures, and check whether the resulting expressive robot could be perceived as coming from one culture or another.

6 Future Trends

The developmental paradigm to the construction of artificial emotion systems creates very promising new areas of research.

Firstly, the practical applications are wide. Consider the development of robot personality, using these flexible definitions of emotion as a basis. We could imagine a robot that could adapt to the personality or emotional style of its family: very expressive and outgoing, or quiet and reserved. A robot could express happiness in a way that is consistent with its surroundings, possibly increasing user acceptance. Or, a system could simply improve its emotion recognition accuracy through interaction with the users – just as we may understand a good friend's true feelings (even when they try to hide it), the system could adapt its definition of emotion by linking together person-specific facial features, vocal features, and context.

Secondly, the developmental paradigm may answer the question: what is robot love? This question has captured the interest of the media, films, and pop culture, and is certainly a valid question for the topic of this book. Consider the GMM link between humans, emotion states, and a robot's physical "gut feelings". If a robot continues to associate physical *flourishing* with not only emotional features, but also physical features (like a caregiver's face), it could develop attachment. This is a fascinating idea that suggests that robot companions could be "loving" agents. For instance, a caregiver's presence could make the robot "happy", associate it with "full battery", and its presence would therefore be akin to repowering itself at a charging station, like the idea that a loved one re-energizes us. As Daniel Dennett supposed when pressed with the question: "In principle, you can make a computer that loved right out of the box, but only because it was a near-duplicate of a computer that had a life, that had love. There's probably no shortcut." (So, 2013)

Thirdly, we should talk about empathy and its links to moral machines. Affective empathy is the idea that, through emotional contagion, we feel others' pain even out of our conscious control. It has links to morality (de Waal, 2013), and Baron-Cohen suggests that personality disorders such as psychopathy and narcissism are linked to a lack of affective empathy (Baron-Cohen, 1996). Since robot morality and ethics is a debated topic, it is possible that a system like MEI could serve as a platform for affective empathy – a building block for robots to develop morality outside of a rule system (Wallach & Allen, 2010).

Fourthly, we can consider more research in the extension of the MEI system to other modalities and emotions. Touch is an important method of communication with infants, and even animals: as noted by one of our participants, “I wanted to comfort the robot by touching him – that’s how I communicate with my dog.” Of course, facial expressions are also extremely important. While considered in our paradigm (Figure 4), they have not yet been tested since our NAO robot did not have a moveable face. Further research with very human-like robots such as Affetto is an obvious next step. In terms of other emotions, we have touched only briefly on other emotions such as fear and anger. Many emotion classification methods treat happiness, sadness, anger and fear at the same level. But it is known that anger requires higher cognitive mental processing (Fellous & Arbib, 2005). Therefore, emotion modeling in a scaffolded hierarchy could also be explored.

Finally, emotion, thought to be one of the first capacities built by a child, should benefit the field of artificial intelligence and epigenetic robots greatly. The constructional link between the physical body, mental states and preference could be used as a scaffold for learning. For example, consider the concept within artificial intelligence called symbol grounding. Words and concepts are typically grounded in visual features like “the apple is round and red”. Yet when we look at an object, we also have an associated feeling–“good” or “bad”–stemming from outside of consciousness, e.g., “the apple is delicious and I like it”. This feeling could be fundamental in constructing meaning. Therefore, our artificial emotion system and “gut feeling” definition could be key to grounding A.I. and understanding.

7 Conclusion

In this chapter we explored the epigentic robotics paradigm for developing artificial emotions. We suggested how emotions could be constructed through environment, given some initial “innate” assumptions about the system and a human-like interaction called motherese. Importantly, for the first time, we described experiments with a robot that developed *feelings*: physical flourishing or distress grounded in battery levels and motor temperatures. Towards robot emotions that are flexible, culture-specific, and grounded in the physical world, with the potential to be empathetic and moral machines, developmental robotics paradigm is an exciting approach for the future of artificial emotion systems.

8 References

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive Developmental Robotics: A Survey. *Transactions on Autonomous Mental Development , 1* (1), 12-32.

Averill, J. R. (1980). A constructivist view of emotion. In R. Harré, *Emotion: Theory, research, and experience* (pp. 305-339). New York: Academic Press.

Baron-Cohen, S. (1996). Is there a normal phase of synaesthesia in development. *Psyche , 2* (27), 223-228.

Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science* , 28-58.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 1). Springer, New York.

Boucenna, S., Gaussier, P., Andry, P., & Hafemeister, L. (2010). Imitation as a Communication Tool for Online Facial Expression Learning and Recognition. *IROS* (pp. 5323-5328). IEEE/RSJ.

Breazeal, C. L. (2004). *Designing Sociable Robots.* MIT press.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery , 2* (2), 121-167.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. *Interspeech*, (pp. 1517-1520).

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognitiion and experience of disgust following brain injury. *Nature neuroscience , 3* (11), 1077-1078.

Calkins, S. D. (2002). Does aversive behavior during toddlerhood matter? The effects of difficult temerament on maternal perceptions and behavior. *Infant Mental Health Journal* , 381-394.

Camras, L. A. (2011). Differentiation, Dynamical Integration and Functional Emotional Development. *Emotion Review* , 138-146.

Camras, L. A., & Shuster, M. M. (2013). Current Emotion Research in Developmental Psychology. *Emotion Review* , 321-329.

Camras, L. A., Lambrect, L., & Michel, G. (1996). Infant "surprise" expressions as coordinative motor structures. *Journal of Nonverbal Behavior , 20*, 183-195.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). 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.

Cornelius, R. R. (2000). Theoretical Approaches to Emotion. Newcastle: ISCA.

Craig, A. D. (2009). How do you feel–now? the anterior insula and human awareness. *Nature Reviews Neuroscience , 10*, 59-70.

Damasio, A. (1994). Descartes' error: Emotion, reason and the human mind. *New York: Grossett/Putnam* .

Damasio, A. R. (2004). *Looking for Spinoza: Joy, sorrow and the feeling brain.* Random House.

Darwin, C. (1872/1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press.

de Waal, F. (2013). *The Bonobo and the Atheist: In Search of Humanism Among the Primates.* New York: W.W. Norton and Co.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion , 6* (3-4), 169-200.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* , 49-98.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science , 12*, 159-164.

Fellous, J.-M., & Arbib, M. A. (2005). *Who needs emotions?: The brain meets the robot.* Oxford University Press Oxford.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child development* , 1497-1510.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology , 20* (1), 104.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of child language , 16* (3), 477-501.

Field, T., Healy, B., Goldstein, S., Perry, S., Bendell, D., Schanberg, S., et al. (1988). Infants of depressed mothers show "depressed" behavior even with nondepressed adults. *Child development* , 1569-1579.

Fries, A. B., & Pollak, S. D. (2004). Emotion understanding in postinstitutionalized Eastern European children. *Development and Psychopathology , 16* (2), 355-369.

Frijda, N. H. (1986). *The emotions.* Cambridge: Cambridge University Press.

Heberlein, A. S., & Atkinson, A. P. (2009). Neuroscientific Evidence for Simulation and Shared Substrates in Emotion Recognition: Beyond Faces. *Emotion Review , 1* (2), 162-177.

Hess, U., Beaupré, M. G., & Cheung, N. (2002). Who to whom and why – cultural differences and similarities in the function of smiles. In M. Abel, *An empirical reflection on the smile* (pp. 187-216). New York: The Edwin Mellen Press.

Hiatt, S., Campos, J., & Emde, R. (1979). Facial patterning and infant emotional expression: Happiness, surprise and fear. *Child Development , 50*, 1020-1035.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience , 7* (12), 942-951.

Ishihara, H., Yoshikawa, Y., & Asada, M. (2011). Realistic child robot “Affetto” for understanding the caregiver-child attachment relationship that guides the child development. *IEEE International Conference on Development and Learning - ICDL. 2*, pp. 1-5. Frankfurt: IEEE.

James, W. (1884). What is an emotion? *Mind , 19*, 188-205.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin , 129* (5), 770.

Kaffman, A., & Meaney, M. J. (2007). Neurodevelopmental sequelae of postnatal maternal care in rodents: Clinical research and implications of molecular insights. *Journal of Child Psychology and Psychiatry , 48*, 224-244.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on , 24* (7), 881-892.

Kim, P., Feldman, R., Mayes, C., Eicher, V., Thompson, N., Leckman, J., et al. (2011). Breastfeeing, brain activation to own infant cry, and maternal sensitivity. *Journal of Child Psychology and Psychiatry , 52* (8), 907-915.

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy , 14* (1), 77-100.

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition & Emotion , 14* (1), 93-124.

Kohler, E. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science , 297* (5582), 846-848.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience , 5* (11), 831-843.

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics & Gynecology and Reproductive Biology , 68*, 1-23.

Lim, A., & Okuno, H. G. (in prep.). Emotional Contagion as a Mechanism for Basic Empathy.

Lim, A., & Okuno, H. G. (in press). The MEI Robot: Towards Using Motherese to Develop Multimodal Emotional Intelligence. *IEEE Transactions on Autonomous Mental Development* .

Lim, A., Ogata, T., & Okuno, H. G. (2011). Converting emotional voice to motion for robot telepresence. *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, (pp. 472-479).

Lim, A., Ogata, T., & Okuno, H. G. (2012). 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.

Mason, W. A., & Capitanio, J. P. (2012). Basic Emotions: A Reconstruction. *Emotion Review* , 238-244.

Meaney, M. J., & Ferguson-Smith, A. C. (2010). Epigenetic regulation of the neural transcriptome: The meaning of the marks. *Nature Neuroscience , 13*, 1313-1318.

Mehrabian, A. (1995). {Framework for a comprehensive description and measurement of emotional states.}. *Genetic, social, and general psychology monographs , 121* (3), 339-361.

Morrison, I., Mjornsdotter, M., & Olausson, H. (2011). Vicarious Responses to Social Touch in Posterior Insular Cortex are Tuned to Pleasant Caressing Speeds. *Journal of Neuroscience , 31* (26), 9554-9562.

Nagy, E., Pilling, K., Orvos, H., & Molnar, P. (2013). Imitation of tongue protrusion in human neonates: Specificity of the response in a large sample. *Developmental psychology , 49* (9), 16-28.

Ortony, A. (1990). *The cognitive structure of emotions.* Cambridge university press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research , 12*, 2825-2830.

Pohl, A., Anders, S., Schulte-Ruther, M., Mathiak, K., & Kircher, T. (2013). Positive Facial Affect – An fMRI Study on the Involvement of Insula and Amygdala. *PLoS ONE , 8* (8), e69886.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience , 27*, 169-192.

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion , 23* (7), 1259-1283.

Sabatini, M. J., Ebert, P., Lewis, D. A., Levitt, P., Cameron, J. L., & Mirnics, K. (2007). Amygdala gene expression correlates of social behavior in monkeys experiencing maternal separation. *Journal of Neuroscience , 27* (12), 3295-3304.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., et al. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition?(A Systematic Review). *PloS one , 8* (10), e78103.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics , 6* (2), 461-464.

Shute, B., & Wheldall, K. (1989). Pitch alterations in British motherese: some preliminary acoustic data. *Journal of Child Language , 16* (3), 503-512.

Sigelman, C. K., & Rider, E. A. (2010). *Life span human development.* CengageBrain.

Singer, T. (2004). Empathy for Pain Involves the Affective but not Sensory Componets of Pain. *Science , 303* (5661), 1157-1162.

So, J. (2013, 12 31). *Can Robots Fall in Love, and Why Would They?* Retrieved 4 7, 2014, from Interview in The Daily Beast: http://www.thedailybeast.com/articles/2013/12/31/can-robots-fall-in-love-and-why-would-they.html

Soken, N. H., & Pick, A. D. (1992). Intermodal Perception of Happy and Angry Expressive Behaviors by Seven-Month-Old Infants. *Child development , 63* (4), 787-795.

Sokol, S. (1978). Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Research , 18* (1), 33-39.

Tottenham, N., Hare, T. A., Millner, A., Gilhooly, T., Zevin, J., & Casey, B. J. (2011). Elevated Amygdala Response to Faces Following Early Deprivation. *Developmental Science , 14* (2), 190-204.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological science , 11* (3), 188-195.

Tsai, J. L. (2007). Ideal Affect: Cultural Causes and Behavioral Consequences. *Perspectives on Psychological Science , 2* (3), 242-259.

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other.* Basic Books.

Wallach, W., & Allen, C. (2010). *Moral Machines: Teaching Robots Right from Wrong.* New York: Oxford University Press.

Watanabe, A., Ogino, M., & Asada, M. (2007). Mapping facial expression to internal states based on intuitive parenting. *Journal of Robotics and Mechatronics , 19* (3), 3-15.

Werner, L. A. (2002). Infant auditory capabilities. *Current Opinion in Otolaryngology & Head and Neck Surgery , 10* (5).

Key Terms and Definitions

- **Bayesian Information Criterion (BIC)**: A criterion for selecting the best model given a dataset. It penalizes models that use too many variables to explain the data (overfitting). A lower BIC implies 1) a better fit, and/or 2) fewer explanatory variables.
- **Developmental Robotics** or **Epigenetic Robotics:** A relatively new scientific field that aims to study the developmental mechanisms and architectures for lifelong learning in machines. Typically it involves formalizing, validating and extending models from neuroscience, developmental psychology, and evolutionary biology, specifically by attempting to implement the models in robots. Results are expected to feedback into existing theories, or produce novel theories about human and animal development.
- **Entrainment:** The synchronization of organism to a rhythm usually produced by another social actor. Humans can entrain to the beat, for instance, by dancing or tapping their foot, and fireflies are also known to flash in synchrony. In this chapter, we refer to the entrainment in speed, intensity, regularity and extent between the voices and movements of two agents.
- **Gaussian Mixture Model (GMM):** A probabilistic model used to represent data as a mixture of normal distributions. It is commonly used for unsupervised learning and clustering, which means that clusters can be created without labels. It is similar to k-means clustering, except that when used for recognition, it outputs the probability that a new data point belongs to a cluster, not a binary value.
- **(Gut or Physical) Feeling:** The state of physical flourishing (homeostasis) or distress (out of homeostasis) in an individual.
- **Motherese:** A simplified type of speech spontaneously spoken by caregivers to infants. Typically it contains exaggerated intonation and rhythm, a higher pitch, and more pronounced variations compared to normal speech. Also known as baby talk or infant-directed speech (IDS).
- **Multimodal Emotional Intelligence (MEI):** A robot system with the ability to understand, represent and express emotions in multiple modalities, such as voice, movement, gait or music.
- **SIRE Paradigm:** A paradigm using speed, intensity, regularity and extent (SIRE) to represent an emotion across modalities. For instance, sadness has been linked to slow, low intensity, regular and small dynamics in movement, as well as in voice and music.
- **Social constructivist theory:** A theory that an individual's learning is constructed through interaction with others in a group. It suggests that cognitive development is influenced by culture and social context.

---

[i] http://www.aldebaran.com
[ii] http://www.hark.jp
[iii] http://www.acapela-group.com