Personalized Robot Emotion Representation through Retrieval of Memories

Thi Le Quyen Dang, Sungmoon Jeong, Nak Young Chong
School of Information Science
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
e-mail: {quyen.dang, jeongsm, nakyoung}@jaist.ac.jp

***Abstract*—We present a robot emotion representation model by investigating the role of human-inspired social effects and memory retrieval process. Social referencing and social sharing are modeled as human guides to share knowledge and direct or influence robot emotion generation. The robot's acquired knowledge is consolidated into a developmental memory architecture and used for future retrieval. This model enables robots interacting with their environment to learn from humans and form a human-like emotion generation process which helps to facilitate personalized human-robot interaction.**

Keywords-human-robot interaction; robot emotion; developmental architecture; social sharing; social referencing

I. INTRODUCTION

Emotion has been known as a key factor to enhance human-human interaction [1] and helps to infer an individual's certain action, as it is often directed toward a specific object. Since humans tend to treat robots by the way they treat others, emotions should be generated appropriately by robots to facilitate natural human-robot interaction [2]. Cognitive robotics researchers investigate the role of memory retrieval process of humans in emotion activation, response difference, and personality formation [3]-[5] to model emotion as a module of memory. Artificial emotions enable robots to remember affective experiences and drive a mechanism for action selection and complete given tasks effectively [6]-[8].

However, emotions of an individual are activated and influenced not only by personal experiences but also social effects due to social referencing and social sharing. On one hand, social referencing, which acts strongly in human infants, helps to teach an individual a basic knowledge about a particular object or stimulus and shape their emotions and behaviors in response to that stimulus [9], [10]. On the other hand, social sharing helps the individual generate emotions and drive emotional responses for a more detailed information about the environment by using the acquired personal knowledge and experiences [11]. Thus, modeling artificial emotions for robots is needed to consider both the roles of past experiences and social effects. A biologically-inspired computational model of social referencing was presented by building an expressive human-like robot, Leonardo [12]. This robot is able to share human attention mechanism, mimic emotional expressions, and interpret the environment by accepting human guides as a passive learner. However, previous works were limited on the simulation of human cognition based on the cognitive memory architecture and ignored the role of social effects from the perspective of sharing.

In this paper, we propose a new approach in representing robot emotions that would be favorably accepted by humans employing a developmental memory architecture and investigating social effects on both aspects of referencing and sharing. To apply a theory of developmental memory architectures, we use Epigenetic Robot Intelligent System (ERIS) [13] in modeling our robot long-term memory (LTM) architecture. Besides, social effects are considered as human guides during human-robot interaction. To represent our robot emotions, the valence-arousal emotional model is used considering their effects on memory and appraisal process applied in an infant development-inspired robot [12], [14]. This representation allows robots to use both the guided social knowledge and personal experiences on generating emotions individually in response to the environmental stimulus.

This paper is organized as follow. Section II presents our proposed robot emotion representation based on past experiences and social effects. Section III describes details of our experiments. Section IV summarizes and analyzes the experiment results obtained. Section V draws conclusions and outlines directions for future research.

II. A ROBOT EMOTION REPRESENTATION BASED ON MULTIPLE EMOTION ACTIVATION CUES

A. *Representation Scale: Valence-Arousal*

In an attempt to classify and represent human emotions, psychologists proposed two main directions: emotions as discrete categories and dimensional models, respectively. Along the lines, a 2-dimensional emotional framework [13] has been used to design a robot emotion model based on the effect of social referencing through examining human emotions with facial expressions and affective states [14]. This framework suggests to distribute emotions as a combination of two values on the valence-arousal coordinates as shown in Fig. 1. While valence shows how negative or positive the emotion is, arousal presents the intensity of that emotion. For example, angry can be described by a combination of negative valence and high arousal values. In addition, valence and arousal have been known as key factors which enhance memory performance through encoding, consolidation [15]-[17], and retrieval [18], [19] processes. Therefore, in our robot emotion representation, we use the framework to investigate and model the role of past experiences and social effects.

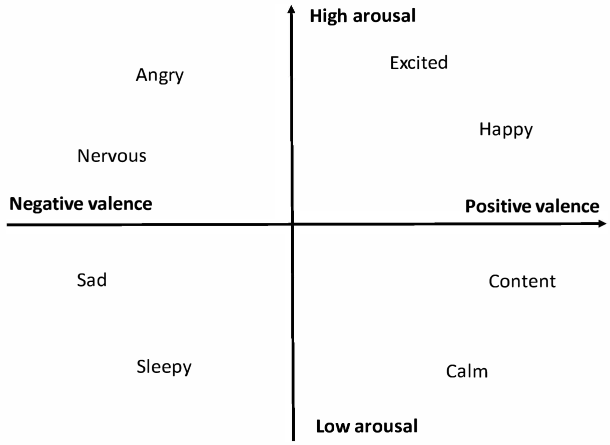

Figure 1. 2-dimentional emotional framework

B. Cue 1: Social Effects

In human-human interactions, social effects are considered by the following two types: social referencing and social sharing. On one hand, social referencing is the fundamental of learning which allows infants to focus on some specific object or stimulus, thereby making themselves adapt to the environment by learning and copying from their parents and caregivers [9], [10]. Through interacting with those guides, infants gain knowledge and create their own appraisal about a novel object or stimulus of the environment with social context. For example, infants increase their distance to their mother when receiving joyful messages, or move closer to their mother when getting fearful messages [10]. On the other hand, social sharing helps to facilitate human relationships and allows subjects to gain a detailed interpretation about the environment based on social context. Children older than 18 months start to negotiate with their guide about the way to understand and generate emotions for object appraisals based on their knowledge [9]. Sharing aspect of social effects takes a stronger role than referencing aspect to older children and adult [20], [21].

To model social effects to be given to robots, we consider human guides for the role of teaching and sharing during human-robot interaction by providing each robot with common knowledge. The guide provides robots with a common understanding of emotions on a specific object so that robots can further create appraisals related to that object. Initially, robots accept a human guide as a passive learner as inspired from social referencing [9]. After several interactions, robots reduce an agreement on human guide directly through interactions, and they consider guided emotions as an affective factor for sharing knowledge about object appraisals to generate emotions. This influence depends on the intensity of the relationship of human and guided emotion components. In this research, we consider that every robot has the same relationship with humans, so we explore the influence only by the intensity of guided emotion components.

Thus, we can represent the robot personal emotion component (C_{PEt}) at time t is updated based on the influence of the same component of the guided emotion (C_{GEt-1}) by the parameter λ given by

$$C_{PEt} = \lambda\, C_{GEt-1} + (1 - \lambda)\, C_{PEt-1}. \quad (1)$$

Here, λ, which describes the intensity of the guided emotion and shows the role of the guided emotion to the robot personal emotion, is calculated as follows:

$$\lambda = \frac{C_{GEt-1}}{C_{GEt-1} + C_{PEt-1}} \quad (2)$$

Specifically, robots learn from humans when λ is equal to 1, and share with humans otherwise.

C. Cue 2: Past Experiences

When recalling past experiences, humans relive the original experience including all engaged sensory, contextual, semantic, and emotions based on the relationship among LTM components [22], [23], leading to the changes in subjects' current emotional state [24]. The high frequency of recalling times increases the recalling possibility of that event in the future and let the subject remember original events more vividly [25]. Preschool children practice remembering negative events and emotions when talking to their parents; then, they may face the same negative effect when remembering about past negative events [5].

In this representation, we assume that only experienced emotions can influence robot personal emotion due to the effects of experienced emotion components such as valence and arousal. At time t, each robot personal emotion component (C_{PEt}) is updated due to a corresponding experienced emotion component (C_{EEt-1}) and intensity of the component based on the parameter β as follows:

$$C_{PEt} = \beta\, C_{EEt-1} + (1 - \beta)\, C_{PEt-1}. \quad (3)$$

Here, β, the influence of the experienced emotion component to the corresponding robot personal emotion component, is calculated as follows:

$$\beta = \frac{C_{EEt-1}}{C_{EEt-1} + C_{PEt-1}} \quad (4)$$

When the robot does not have any past experiences, there is no experienced emotion, so β is assigned to 0.

D. Long-Term Memory Architecture based on ERIS

In order to enable robots to gain knowledge and remember affective appraisal of objects from human guides for future use, we provide robots with Visual Feature Extractor function. Robots are able to extract visual features of objects in response to a specific visual stimulus to encode an episode. Besides, the robot personal emotion is updated through Emotion Generation Processor function. The episode then is integrated with the updated emotion to be consolidated into LTM of the robot for object affective appraisal formation and future retrieval detailed in Fig. 2.

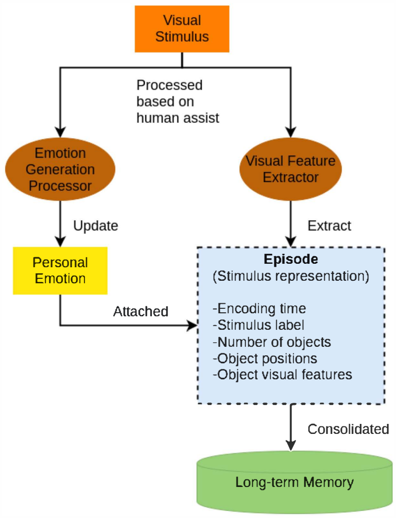

Figure 2. A memory consodilation process.

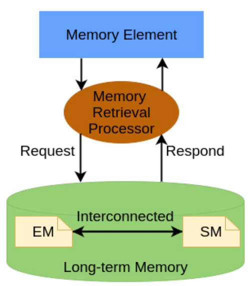

Figure 3. A memory retrieval process.

Our LTM, which is designed based on ERIS architecture, consists of two interconnected components: Episodic Memory (EM) and Semantic Memory (SM). While EM maintains visual stimuli representations including engaged emotions as episodes, SM represents acquired novel knowledge about object appraisals such as color and shape features. The interconnection of SM and EM let the robots retrieve experienced emotions through the similarity of current object and old objects during the memory retrieval process as shown in Fig. 3.

We define the similarity between two objects by comparing object visual features: color and shape [26] of the new object (O_n) to a previously seen object (O_s). The Euclidian distance is used for the calculation of the difference between color and shape features of two objects, where the difference is represented as d_{color} and d_{shape}. The role of each difference on the distance calculation is defined as the parameter ε given by

$$d(O_n, O_s) = \varepsilon.d_{color}(O_n, O_s) + (1 - \varepsilon).d_{shape}(O_n, O_s) \quad (5)$$

From the calculated distance, we get the smallest distance which shows the most similar object previously seen to the new object.

E. Integrating Multiple Cues in Robot Emotion Generation

Our robot emotion representation aims at integrating the influence of human guides based on social effects and robot personal experiences through the memory retrieval process, generating the robot personal emotion due to time parameter and interactions with humans. The overview of this emotion generation process is shown in Fig. 4.

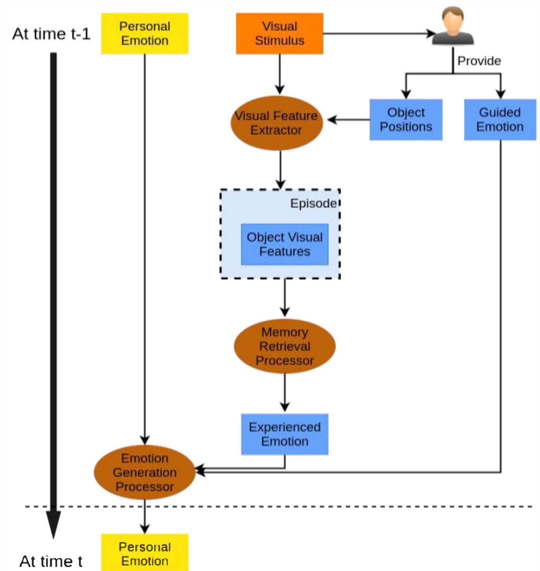

Figure 4. Robot personal emotion generation process.

For every interaction with humans, robots are provided with the location information of all objects in a visual stimulus and a guided emotion. On the one hand, the provided location of objects helps robots identify and extract visual features of those objects. After that, they are able to obtain their past experience including engaged emotion which is an experienced emotion through Memory Retrieval Processor. On the other hand, the guided emotion directs and shapes robot personal emotions to be accepted by humans. Based on the experienced emotion and the guided emotion, the robot personal emotion is updated through Emotion Generation Processor (EGP) function.

In order to build EGP function for updating robot personal emotion components, we apply the same method designed (1), (2), (3) and (4) as follows:

$$C_{PEt} = \gamma_1 C_{GEt\text{-}1} + \gamma_2 C_{EEt\text{-}1} + (1 - \gamma_1 - \gamma_2) C_{PEt\text{-}1} \quad (6)$$

where γ_1 and γ_2 are calculated as follows, respectively:

$$\gamma_1 = \frac{C_{GEt\text{-}1}}{C_{GEt\text{-}1} + C_{EEt\text{-}1} + C_{PEt\text{-}1}} \quad (7)$$

$$\gamma_2 = \frac{C_{EEt\text{-}1}}{C_{GEt\text{-}1} + C_{EEt\text{-}1} + C_{PEt\text{-}1}} \quad (8)$$

Our robot emotions are represented by valence and arousal, but those components do not impact the same way to the subject. While only high arousal works on directing robot attentions from humans and enhancing robot memory performance, both the negative and positive valences do similar influences. Hence, we use a parameter, μ, to present the lowest intensity of each component and update (7) and (8) as follows:

$$\gamma_1 = \frac{|C_{GEt\text{-}1} - \mu|}{|C_{GEt\text{-}1} - \mu| + |C_{EEt\text{-}1} - \mu| + |C_{PEt\text{-}1} - \mu|} \quad (9)$$

$$\gamma_2 = \frac{|C_{EEt\text{-}1} - \mu|}{|C_{GEt\text{-}1} - \mu| + |C_{EEt\text{-}1} - \mu| + |C_{PEt\text{-}1} - \mu|} \quad (10)$$

After being updated, the personal emotion of the robot is encoded into the same place in LTM with the encoded visual features of the stimulus through memory consolidation processes as explained in Section II.D.

III. Experiment

A. Dataset

We use International Affective Picture System (IAPS) [27] and Nencki Affective Picture System (NAPS) [28] datasets. Those datasets are standard emotion evoking image sets in psychology. They contain natural color images in several categories of humans, animals, food, objects, and landscapes. The rating of each image is associated with a mean and a standard deviation of valence and arousal by human evaluation. Valence and arousal are ranged from 1 to 9. Non-arousing neutral emotion is represented as (5, 1) point on the valence-arousal coordinates.

In this research, we use only cat images with simple backgrounds, so we choose 12 images from two datasets with valence (mean: 5.365, std: 3.525) and arousal (mean: 5.095, std: 2.405). All images are rescaled to the same size. Those images also contain only one object, so we assume that the rating of humans for each image is the rated emotion for the contained object. The rating values of valence and arousal are limited; we selected images based on four types such as negative-low arousing, negative-high arousing, positive-low arousing, and neutral-mid arousing. Fig. 5 shows some example of selected images.

Figure 5. Samples of selected images (a) neutral-mid arousing image, (b) and (c) positive-low arousing image.

B. Demo Scenario

To understand how the proposed system works, we set up a scenario in detail, where humans provide guided emotions of every input image to robots. Images are given to a robot in a certain order. In this scenario, the robot plays a role as a passive learner and represents emotions by showing the values of valence and arousal.

Firstly, the robot receives a human guide such as guided emotional values and object localization. The robot then extracts visual features of the images and contained objects in a localized area, then this information is used for recalling similar past experiences and consolidating a new episode. The robot uses extracted visual features of localized objects such as color and shape to look for the past experience of those features and recall experienced emotions.

Secondly, the robot updates personal emotional components based on the corresponding value of experienced emotion and guided emotion and (6).

Lastly, generated personal emotional values are attached to extracted visual features to an episode to be consolidated into the robot LTM and used for future processes. Note that we do not apply any forgetting mechanism based on the assumption made in ERIS [13].

C. Experiment Setup and Procedure

The experimental procedure contains two main implementations. The first one is to investigate the role of social effects on robot emotion generation from the two aspects of social referencing and social sharing. In this experiment, we expect that the robot generates emotions for a certain image differently if it is observed at different time instants.

The second experiment is performed to explore robot emotion representations by both the influence of social effects and past experiences. The purpose of this implementation is to investigate the difference of each robot in emotion generation process for the same stimulus based on the same applied emotion representation system and human guides.

For each experiment, humans provide robots with several images in a certain order, robots then display emotional values of personal emotions, guided emotions, and experienced emotions if it has similar experiences before.

D. Experiment 1: Investigating Social Effects

In this experiment, one robot is used to generate emotions based on visual stimulus and human guides. To do this experiment, we set the robot emotion to neutral non-arousing (5, 1) initially.

We use five different images which include one negative-low arousing image, one negative-high arousing image, two positive-low arousing images, and one neutral-mid arousing image, respectively. These images which are numbered from 1 to 5 are shown to the robot 10 times according to the order of 5, 1, 2, 5, 3, 4, 5, 1, 4, and 5.

E. Experiment 2: Robot Emotion Generation based on Multiple Activation Cues

Two robots are used in this experiment and set to have the neutral-non arousing emotional state (5, 1) initially. Besides, two robots are provided with the same human guides.

Eight different images, numbered from 1 to 8, are shown which contain one negative-low arousing image, one negative-high arousing image, four positive-low arousing images, and three neutral-mid arousing images, respectively. The first robot observes the 8 images in the sequence of 1, 2, 3, 4, 5, 6, 7, and 8. The second robot observes them in the sequence of 2, 3, 1, 4, 6, 7, 5, and 8. In those sequences, two robots see image 4 and image 8 at the same time.

IV. Results and Discussion

A. Results

The result of the first experiment is given in Fig. 6 showing the valence value and arousal value of the robot personal emotion and guided emotion. Fig. 7 shows only the personal emotion components of two robots in the second experiment. We obtained those results by setting ε equal to 0.5 in (5), and μ equal to 1 and 5 for updating valence and arousal, respectively, in (9) and (10).

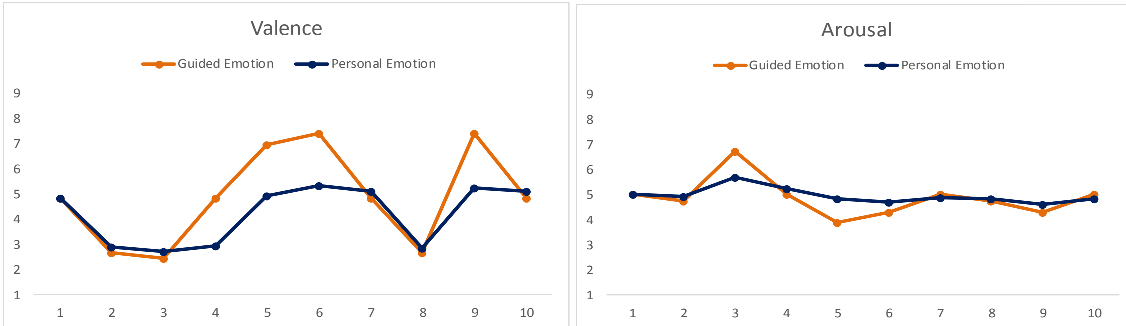

Figure 6. Personal emotions and guided emotions of the robot in the first experiment.

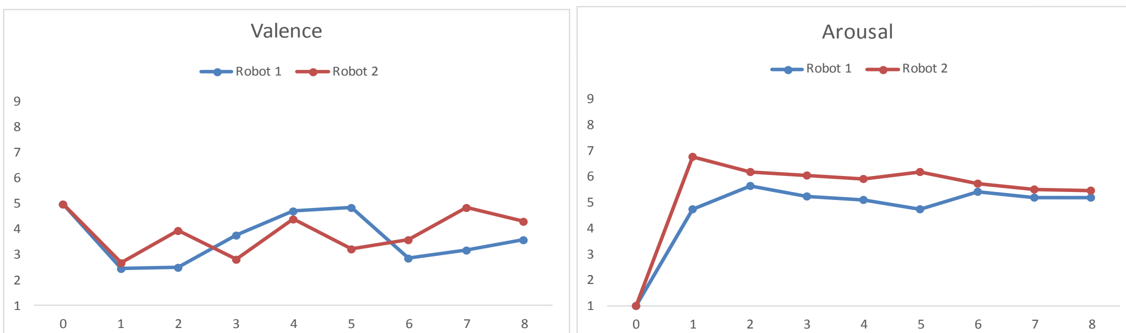

Figure 7. Personal emotions of two robots in the second experiment.

B. Discussion

For the first experiment, when the robot sees the image 5 during the first interaction, the robot generates the same values of valence and arousal by the guided emotion. It is similar to human infants when they learn and mimic the way their parent generates emotions directly. However, the robot does not generate the same emotional values for the images 1, 2, 3, and 4 when it sees those images firstly at the interaction number 2, 3, 5, and 6, respectively. Besides, the emotional values of the robot at the interaction number 4, 7, and 10 are different from the first interaction and guided emotions, when seeing the same image 5. This effect happens similarly for the first time and the second time the robot sees the images 1 and 4. It can be explained with the change of the role of human guide to the robot emotion generation process. After the first interaction, the robot gained a specific knowledge about cats and create affective appraisal to the object by extracting visual images based on human guides. From the second interaction, the robot starts to reduce the consideration about guided emotions which humans provide. Humans start to take a role of sharing with the robot for influencing robot emotion generation direction. Therefore, both roles of humans toward robots, which are teaching and sharing, are modeled based on human-inspired social effects to guide and effect the way robot understand and generate emotions in response to the environmental stimuli.

In the second experiment results, the personal emotions of two robots are presented each time those robots see an image. Although two robots see the same images, initiated by the same emotion state and provided by the same human guided emotions, they generate different emotions for the same image such as the images 4 and 8 at the same time. All images given to both robots are the same but they are arranged in different orders. That leads to forming different gained knowledge and generated emotions about each object appraisal. When the retrieval process occurs, the robot recalled different experienced emotions for the same object which is mostly similar to the new object to generate emotions. Besides, personal emotions of two robots at the time when encoding the images 4 and 8 are different from each other. Different past experiences and different personal emotions at the time when encoding a new experience enable robots to generate different emotions for the same stimulus, which may lead to creating the robot personality.

Our two experiments show that the proposed emotion representation provides robots with a capability to generate emotions similar to the way humans do based on social effects and past experiences. Two different robots are able to have different emotions for the same stimulus due to their own personal knowledge which is taught and shared by humans.

V. Conclusion and Future Works

In this paper, we present a robot emotion representation which is inspired by social psychology and developmental psychology to provide robots with a capability to generate emotions and develop in a human-like way. Social effects are investigated and implemented to personalize robots from both the social referencing and social sharing perspectives. Human plays a role as a teacher and a partner for guiding and sharing emotions and knowledge to direct or influence robot emotion generation or further facilitate a sustainable human-robot interaction.

Emotion generation in this representation is investigated based on object affective appraisals with a similar mechanism as human infants, while emotions of children and adults are generated mainly by situation awareness ability. In the future, we will elaborate on human cognitive skills and abilities for recognizing environmental situations, and implement those features on a real robot.

Acknowledgment

This project was supported by the EU-Japan coordinated R&D project on "Culture Aware Robots and Environmental Sensor System for Elderly Support" commissioned by the Ministry of Internal Affairs and Communications of Japan and EC Horizon 2020.

References

[1] Jaimes, Alejandro, and Nicu Sebe. "Multimodal human–computer interaction: A survey." *Computer vision and image understanding* 108.1 (2007): 116-134.

[2] Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots." *Robotics and autonomous systems* 42.3 (2003): 143-166.

[3] A. Damasio, The Feeling of What Happens (Harcourt Brace, New York, 1999).

[4] Harris, Paul L., and Mark S. Lipian. "Understanding emotion and experiencing emotion." Children's understanding of emotion (1989): 241-258.

[5] Lagattuta, Kristin Hansen, and Henry M. Wellman. "Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion." Child Development 72.1 (2001): 82-102.

[6] Dodd, Will, and Ridelto Gutierrez. "The role of episodic memory and emotion in a cognitive robot." Proceedings of 14th Annual IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN). 2005.

[7] Leconte, Francis, François Ferland, and François Michaud. "Design and integration of a spatio-temporal memory with emotional influences to categorize and recall the experiences of an autonomous mobile robot." Autonomous Robots 40.5 (2016): 831-848.

[8] Kasap, Zerrin, and Nadia Magnenat-Thalmann. "Towards episodic memory-based long-term affective interaction with a human-like robot." 19th International Symposium in Robot and Human Interactive Communication. IEEE, 2010.

[9] Feinman, Saul. "Social referencing in infancy." Merrill-Palmer Quarterly (1982-) (1982): 445-470.

[10] Klinnert, Mary D. "The regulation of infant behavior by maternal facial expression." *Infant Behavior and Development* 7.4 (1984): 447-465.

[11] Hareli, Shlomo, and Anat Rafaeli. "Emotion cycles: On the social influence of emotion in organizations." *Research in organizational behavior* 28 (2008): 35-59.

[12] Thomaz, Andrea Lockerd, Matt Berlin, and Cynthia Breazeal. "An embodied computational model of social referencing." ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.. IEEE, 2005.

[13] Pratama, Ferdian, et al. "Long-term knowledge acquisition in a memory-based epigenetic robot architecture for verbal interaction." *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*. IEEE, 2015.

[14] Russell JA. A circumplex model of affect. Journal of Personality and Social Psychology. 1980;39:1161–1178

[15] Cahill, Larry, and James L. McGaugh. "A novel demonstration of enhanced memory associated with emotional arousal." Consciousness and cognition 4.4 (1995): 410-421.

[16] Kensinger, Elizabeth A. "Remembering emotional experiences: The contribution of valence and arousal." Reviews in the Neurosciences 15.4 (2004): 241-252.

[17] LaBar, Kevin S., and Elizabeth A. Phelps. "Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans." Psychological Science 9.6 (1998): 490-493.

[18] Bradley, Margaret M., et al. "Remembering pictures: pleasure and arousal in memory." Journal of experimental psychology: Learning, Memory, and Cognition 18.2 (1992): 379.

[19] Gomes, Carlos FA, Charles J. Brainerd, and Lilian M. Stein. "Effects of emotional valence and arousal on recollective and nonrecollective recall." Journal of Experimental Psychology: Learning, Memory, and Cognition 39.3 (2013): 663

[20] Schachter, Stanley, and Jerome Singer. "Cognitive, social, and physiological determinants of emotional state." Psychological review 69.5 (1962): 379.

[21] Mikulincer, Mario, and Phillip R. Shaver. "Attachment theory and emotions in close relationships: Exploring the attachment‐related dynamics of emotional reactions to relational events." Personal Relationships 12.2 (2005): 149-168.

[22] Talarico, Jennifer M., Kevin S. LaBar, and David C. Rubin. "Emotional intensity predicts autobiographical memory experience." Memory & cognition 32.7 (2004): 1118-1132.

[23] Conway, Martin A. "Sensory–perceptual episodic memory and its context: Autobiographical memory." Philosophical Transactions of the Royal Society B: Biological Sciences 356.1413 (2001): 1375-1384.

[24] Ahn, Hyeon Min, et al. "The effect of cognitive reappraisal on long‐term emotional experience and emotional memory." Journal of neuropsychology 9.1 (2015): 64-76.

[25] Lane, Richard D., et al. "Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science." Behavioral and Brain Sciences 38 (2015): e1.

[26] Berryhill, Marian E., and Ingrid R. Olson. "Is the posterior parietal lobe involved in working memory retrieval?: Evidence from patients with bilateral parietal lobe damage." *Neuropsychologia* 46.7 (2008): 1775-1786.

[27] Lang, Peter J., Margaret M. Bradley, and Bruce N. Cuthbert. "International affective picture system (IAPS): Affective ratings of pictures and instruction manual." *Technical report A-8* (2008).

[28] Marchewka, Artur, et al. "The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database." *Behavior research methods* 46.2 (2014): 596-610.