

Affective Dialogue Communication System with Emotional Memories for Humanoid Robots

M. S. Ryoo*, Yong-ho Seo, Hye-Won Jung, and H. S. Yang

Artificial Intelligence and Media Laboratory
Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
{mryoo, yhseo, hwjung, hsyang}@kaist.ac.kr

Abstract. Memories are vital in human interactions. To interact sociably with a human, a robot should not only recognize and express emotions like a human, but also share emotional experience with humans. We present an affective human-robot communication system for a humanoid robot, AMI, which we designed to enable high-level communication with a human through dialogue. AMI communicates with humans by preserving emotional memories of users and topics, and it naturally engages in dialogue with humans. Humans therefore perceive AMI to be more human-like and friendly. Thus, interaction between AMI and humans is enhanced.

1 Introduction

Affective computing is becoming a more important research issue in the area of robotics. Several studies in robotics have been conducted to enable affective interaction between humans and robots [1,2,3]. However, most studies have focused only on primitive interactions, such as facial expressions or body movements. A sociable robot that can naturally interact with humans must be capable of high-level interaction; that is, it must be able to communicate with humans in the form of dialogue. Thus, the robot must be able to synthesize its emotion and express it in a dialogue. Our research enables robots to have high-level dialogue similar to the conversational robot Mel, who attempted to converse with humans [4]. However, Mel ignored the emotions and multimodality of human-robot communication. Our goal is to design and implement a sociable robot that can affectively interact with humans through dialogue.

The major improvement in our framework is the construction of a memory system that stores explicit emotional memories of past events. The literature from cognitive science and neuroscience suggests that emotional memories are vital when the human brain synthesizes emotions [5]. While previous research on sociable robots either ignores the emotional memory or maintains the emotional memory implicitly in high-

* Current address: CVRC, University of Texas at Austin. mryoo@mail.utexas.edu

level interaction, we need to establish explicit memories of emotions, events, and concepts. We have therefore adopted the concept of emotional memory for our humanoid robot. Our memory system maintains explicit memories of previous emotional events. Thus, the emotional system can synthesize emotions on the basis of emotional memories.

In this project, we have developed a general affective dialogue system for a social humanoid robot named AMI. To enable affective dialogue interactions, we constructed an emotional memory system for our framework. The memory system comprises five subsystems: namely, perception, motivation, memory, behavior, and expression. In the perception system, we implemented a bimodal emotion recognizer for recognizing emotions. To ensure AMI can respond appropriately to the emotional status of users and itself, we designed subsystems that use their own drive, emotions, and memory.

2 System Overview

Motivated the human brain structure discovered by cognitive scientist [5], we have designed the framework for our sociable humanoid robot. We designed the affective communication framework to include the five subsystems shown in Fig. 1. Our framework is similar to the creature kernel framework for synthetic characters [6]. The same framework was also applied to the software architecture of Kismet [1]. However, since our goal is to enable dialogue interactions, we improved the framework so that our robot can preserve explicit memories. Thus, our system has two major benefits over older systems.

Memory System. We added a memory system to the referred framework. The memory system enables AMI to represent, and reflect upon, itself and its human partners. It also enhances AMI's social skills and fosters communication with humans. To enable affective interaction between AMI and humans, we enabled AMI to preserve its emotional memory of users and topics.

Dialogue Interaction. Other affective robots, such as Kismet, were based on an infant-caretaker interaction, but our system is based mainly on a dialogue interaction. Accordingly, our internal design and implementation differ from other robots because of our distinct goal of multimodal affective communication.

The main functions of each subsystem are summarized as follows. The perception system, which mainly extracts information from the outside world, comprises the subsystems of face detection, face recognition, emotion recognition, and motion and color detection. The motivation system is composed of a drive and an emotional system. Drives are motivators; they include endogenous drives and externally induced desires. The emotional system synthesizes AMI's artificial emotions. The memory system, as mentioned above, preserves the emotional memories of users and topics. We improved the subsystems of our previous humanoid robots [7].

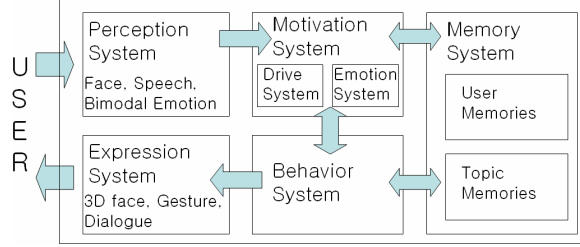


Fig. 1. The figure illustrates the overall framework of the affective system that enables dialogue interactions

3. Perception System

Face Detection and Recognition: The face detection system uses a bottom-up, feature-based approach. The system searches input image for a set of facial features, and groups regions into face candidates based on the geometric relationship of the facial features. Finally, the system decides whether the candidate region is a face by locating eyes in the eye region of a candidate's face. The detected image is sent to the face recognizer and to the emotion recognizer. Face recognizer determines the user's identity from face database. See [7] for more detailed information on implementation.

Bimodal emotion recognition: We estimate emotion through facial expression and speech, and then integrate them to enable bimodal emotion recognition. For emotion recognition through facial expression, we normalized the image captured. We then extracted the following two features, which are based on Ekman's facial expression features [8]: One is Facial image of lips, brow and forehead, and the other is edge image of lips, brow and forehead.

Next, for emotion recognition through speech, we adopted a recognition method similar to the one used in the life-like communication agents MUSE and MIC [9]. Our system extracts two features: a phonetic feature and a prosodic feature. We used a neural network to train each feature vector.

For bimodal emotion recognition, we used decision logic to integrate the two training results. The final result vector of the decision logic (R_{final}) is as follows:

$$R_{final} = (R_{face}W_f + R_{speech}W_s) + R_{final-1} - \delta t. \quad (1)$$

R_{face} and R_{speech} are the results vector of the emotion recognition through facial expression and speech. W_f and W_s are the weights of the two modalities. $R_{final-1}$ is the previous emotion result determined by decision logic, and δ is a decay term that eventually restores the emotional status to neutral. The overall correctness of bimodal emotion system recognition was about 80 percent for each of the five testers. By resolving confusion, the bimodal emotion system performed better than facial-only and speech-only systems.

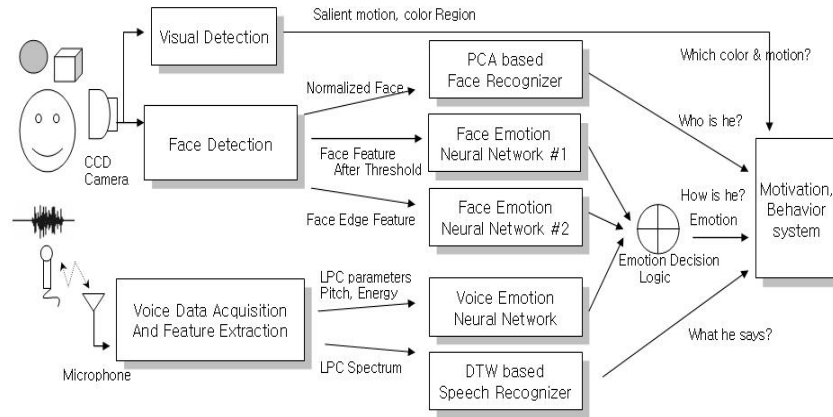


Fig. 2. The figure shows the structure of the perception system

4. Motivation System

4.1 Drive System

Previously, we defined three basic drives for a robot's affective communication with humans: the drive to interact with humans, the drive to ingratiate itself with humans, and the drive to maintain its well-being [7].

The first drive motivates AMI to approach and greet humans. The second drive prompts AMI to make humans feel better. When AMI interacts with humans, it tries to ingratiate itself while considering the human's emotional state. The third drive is related to AMI's maintenance of its own well-being, especially with regard to psychological and physical fatigue. When AMI experiences extreme anger or sadness, or when its battery is too low, it stops interacting with humans.

4.2 Emotional System

Emotions are significant in human behavior, communication and interaction [10]. A synthesized emotion influences the behavior system and the drive system as a control mechanism. To enable AMI to synthesize emotions, we used a model that comprises the three dimensions of emotion [11]. This model characterizes emotions in terms of stance (open/close), valence (negative/positive) and arousal (low/high). Our system always assumes the stance to be open, because AMI is always openly involved in interactions. Therefore, we only consider valence and arousal, implying that only three emotions are possible for AMI: happiness, sadness, and anger.

Arousal factor (CurrentUserArousal): The arousal factor is determined by factors such as whether AMI finds the human, and whether the human responds. Low arousal increases the emotion of sadness.

Valence factor (CurrentUserResponse): The valence factor is determined by whether the human responds appropriately to AMI's requests. A negative response increases the emotion of anger; a positive response increases the emotion of happiness.

The synthesized emotion is also influenced by the drive and the memory system. We used the following equation to compute AMI's emotional status ($E_i(t)$):

$$\begin{aligned} \text{If } t = 0, E_i(t) &= M_i \quad (t = 0 \text{ when new face appears}) \\ \text{If } t \neq 0, E_i(t) &= A_i(t) + E_i(t-1) + \sum D_i(t) + M_i - \delta t. \end{aligned} \quad (2)$$

E_i is AMI's emotional status; t is time; i is happiness, sadness, or anger; A_i is the emotional status calculated by the mapping function of [Arousal, Valence, Stance] from the current activated behavior; D_i is the emotional status defined by the activation and intensity of the unsatisfied drives in the drive system; M_i is the emotion from the memory system; and δt is a decay term that restores the emotional status to neutral.

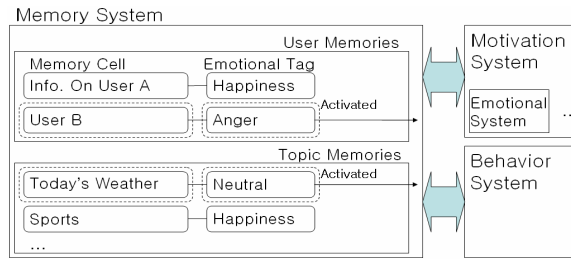


Fig. 3. Activation of memory cells in the memory system

5. Memory System

Our system presents the emotional memories required for more natural and intelligent affective communication with humans. In view of our research objectives, we implemented two memory groups: a user memory, which represents memories of previous interactions with users; and a topic memory, which represents AMI's knowledge of specific topics. When the memory cell is created, AMI's current emotional state is attached to the memory cell as an emotional tag. Later, the memory cells can be activated by AMI's perception system or emotional state. Even though many memory cells can be activated by the perception system or the emotional system, only one memory cell is chosen from AMI's memory group to interact with the other subsystems. The selected memory cell is sent to the behavior system, and the behavior system changes behavior accordingly. Furthermore, the emotional tag is sent to the emotional system as the M_i parameter of the emotional system.

5.1 User Memory

User memories are memories that store information on users. User memories contain data on the user's personality and preference, and this data helps AMI to determine behaviors. Furthermore, as mentioned above, an emotional tag is attached to each memory cell, every time when the memory cell is generated.

After saving data on the user's active and passive responses, AMI uses this data to determine the user's personality. AMI also save the user's preferences, such as likes and dislikes regarding several dialogue topics. This information helps AMI interact appropriately. If AMI recalls this information in the next interaction with the same user, the user thinks AMI is intelligent and the interaction may become more profuse and interesting.

In addition, AMI saves its most frequently occurring emotion, generated by the emotional system mentioned above, as an emotional tag for the latest interaction with the user. These tags subsequently influence AMI's emotional status when it meets the user again. The emotional tags attached to the memory cells are considered AMI's emotional feelings towards the user. Humans often have feelings towards those with whom they communicate. When they meet someone again, they might be influenced by their feelings and memories of the person. The emotional tags in our system are treated similarly: When a memory cell is activated, its emotional tag is passed to the emotional system to help synthesize emotions.

The activation of user memories is represented as follows:

$$U_i(t) = W_{mc} \sum E_k(t) EU_i(t) + W_{mc} C_i. \quad (3)$$

U_i is the activation value of that user memory, t is time, $E_k(t)$ is AMI's current emotion, and $EU_i(t)$ is the emotional tag of the user memory cell. Thus, $\sum E_k(t) EU_i(t)$ indicates the extent of the match between AMI's current emotion and the emotion of the user memory. In addition, C_i is 1 if the current user equals the user of memory U_i . Otherwise, C_i is 0. Finally, W_{me} and W_{mc} are the weight factors.

The memory system selects one user memory after considering the activation values from the perception and emotional systems. That memory is then passed to other systems. Because $U_i(t)$ does not always equal C_i , AMI can activate the memory of a different user with whom it is currently interacting, based on its current emotions.

5.2 Topic Memory

Topic memories contain conversational sentences that AMI has learned from users. The topic memories are first created when the perception system recognizes that the frequency of a keyword has exceeded a threshold; that is, when the user has mentioned the same keyword several times. After AMI's behavior system confirms that the user is talking about a particular keyword, the memory system makes a new topic memory cell for that keyword. In the memory cell, the sentences of the user are stored and an emotional tag is attached with respect to AMI's current emotion.

Of all the topic memories, only the one with the highest activation value is selected at time t . We calculated the activation values of the topic memories, $T_i(t)$, as follows:

$$\begin{aligned} \text{If } \text{COMM} = 0, T_i(t) &= W_{mt} \sum E_k(t) ET_i(t) \\ \text{If } \text{COMM} = i, T_i(t) &= 1. \end{aligned} \quad (4)$$

COMM represents the user's command to retrieve specific topic memory, t is time, $E_k(t)$ is AMI's current emotion, and $ET_i(t)$ is the emotional tag of the topic. Thus, $\sum E_k(t) ET_i(t)$ indicates the extent of the match between AMI's current emotion and the emotion of the memory of the topic. Finally, W_{mt} is a weight factor.

6. Behavior System

Previously, we reported that the structure of the behavior system has three levels, which address the three drives of the motivation system [7]. As the system moves down a level, more specific behavior is determined according to the affective relationship between the robot and human.

The first level of the behavior system is called drive selection. The behavior group of this level communicates with the motivation system and determines which of the three basic drives should be addressed. The second level, called high-level behavior selection, decides which high-level behavior should be adopted in relation to the perception and internal information in the determined drive. In the third level, called low-level behavior selection, each low-level type of behavior is composed of dialogue and gestures, and is executed in the expression system. A low-level type of behavior is therefore selected after considering the emotion and memory from other systems.

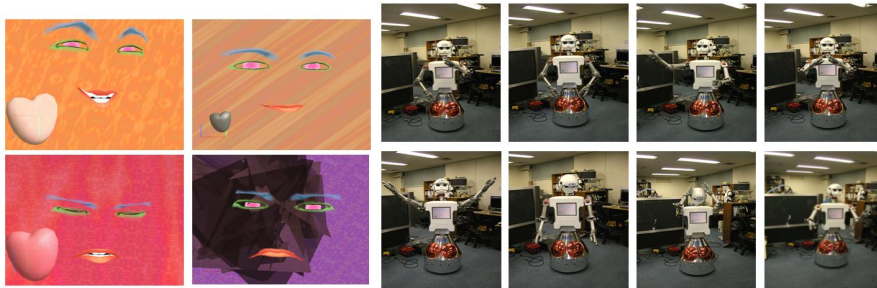


Fig. 4. AMI's facial expressions and gestures

7. Expression System

The expression system, the implementation of which is described in [7], is the intermediate interface between the behavior system and AMI's hardware and it has three subsystems: a dialogue expression system, a 3-D facial expression system, and a gesture expression system.

8. Experimental Results and Conclusions

We first confirmed that each subsystem satisfies its objectives. From our evaluation, we drew the graph in Fig. 5, which shows the subsystem's flow during a sample interaction. The graph also shows the behavior system (finding, greeting, consoling and so on), the motivation system (AMI's drives and emotions), and the perception system (the user's emotional status).

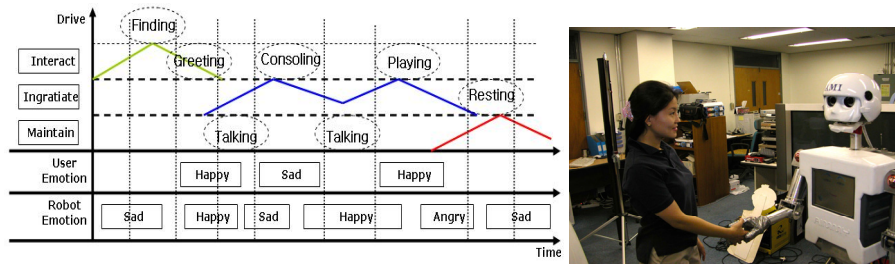


Fig. 5. The work flow of the system

Five users participated in the experiment. We used the procedure of Strack, Martin, and Stepper [12]. The procedure requires participants to hold a pen in their mouths to trigger an emotional response. Each participant was asked to interact with AMI five times while displaying an emotional state of happiness, sadness, or neutrality. For each interaction, all participants evaluated the naturalness of the conversation. They were given three buttons to indicate a negative evaluation: one button indicated an emotional mismatch, another indicated redundancy in the conversation, and the third indicated other errors. If no buttons were pressed, we assumed that particular conversation was natural. Visual input and speech input were given to AMI directly, and all processes were performed on-line.

To evaluate our framework of the emotional memory, we compared three types of systems: one without an emotional or memory system, one with an emotional system but without a memory system, and one with both an emotional and memory system. Table 1 shows the results. The results suggest that the overall performance of the systems with an emotional memory is better than the system without it. The results clearly suggest that emotional memory helps AMI to synthesize more natural emotions and to reduce redundancy in conversational topics.

We have presented an affective communication robot with explicit emotional memories of users and topics. In our system, the emotional memory is the key to improving affective interaction between humans and robots. Other sociable robots have either ignored emotional memories or maintained them implicitly. Our research suggests that explicit emotional memory enhances high-level affective dialogue.

We plan to extend AMI's memory system with a greater variety of memories, including visual objects or high-level concepts. AMI's current memory cells are limited to conversational topics and user-information. In future, AMI will be capable of memorizing information on visual input and word segments, and of making connections between them.

Table 1. Experimental results of interactions

		W/o emotion, w/o memory	With emotion, w/o memory	With emotion, with memory
Natural		54%	70%	76%
Unnatural	Emotion mismatch	19%	10%	5%
	Redundancy	14%	6%	3%
	Other errors	13%	14%	16%

Acknowledgement

This research was partly supported by the Korea Ministry of Science and Technology through the Brain Science Research Center and by the Information Technology Research Center (ITRC) Support Program, the Ministry of Information and Communication.

Reference

1. Breazeal, C.: Designing Sociable Robots. MIT Press, Cambridge, MA (2002)
2. Arkin, R.C., Fujita, M., Takagi, T., Hasegawa, R.: An Ethological and Emotional Basis for Human-Robot Interaction. *Robotics and Autonomous Systems*, 42 (2003)
3. Shibata, T. et al.: Emergence of emotional behavior through physical interaction between human and artificial emotional creature. *ICRA (2000)* 2868-2873
4. Sidner, C.L., Lee, C., Lesh, N.: The Role of Dialog in Human Robot Interaction. *International Workshop on Language Understanding and Agents for Real World Interaction (2003)*
5. Ledoux, J.: *The Emotional brain: the mysterious underpinning of emotional life*. New York: Simon & Schuster (1996)
6. Yoon, S.Y., Burke, R.C., Blumberg, B.M., Schneider, G.E.: Interactive Training for Synthetic Characters. *AAAI (2000)*
7. Jung, H., Seo, Y., Ryoo, M.S., Yang, H.S.: Affective communication system with multimodality for the humanoid robot AMI. *Humanoids (2004)*
8. Ekman, P., Friesen, W.V.: *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, Palo Alto, CA (1978)
9. Tosa, N., Nakatsu, R.: Life-like Communication Agent - Emotion Sensing Character "MIC" & Feeling Session Character "MUSE." *ICMCS (1996)*
10. Armon-Jones, C.: The social functions of emotions. R. Harre (ed.), *The Social Construction of Emotions*, Basil Blackwell, Oxford (1985)
11. Schlossberg, H.: Three dimensions of emotion. *Psychology Review* 61 (1954)
12. Strack, F., Martin, L.L., Stepper, S.: Inhibiting and facilitating condition of facial expressions: A non-obtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54 (1988) 768-777