

Affective Communication Model with Multimodality for Humanoids

Hyun Seung Yang, Yong-Ho Seo, Il-Woong Jeong and Ju-Ho Lee
AIM lab, EECS, KAIST (Korea Advanced Institute of Science and Technology)
Republic of Korea

1. Introduction

A humanoid robot that can naturally interact with humans must be capable of high-level interaction skills; that is, it must be able to communicate with humans in the form of human-like skills such as gestures, dialogue communication with mutual sympathy.

In terms of a human symbiotic situation, it is necessary for a robot to be sociable in terms of thinking and feeling. In this aspect, social interaction and communication that share emotions are important issues in the field of a humanoid robot research.

We propose an affective communication model for humanoid robots, so that these robots achieve high level communication through the framework of the affective communication model with multimodality.

The proposed model comprises five subsystems: namely, perception, motivation, memory, behavior, and expression. In the perception system, we implemented a bimodal emotion recognizer. To ensure a humanoid robot can respond appropriately to the emotional status of users and itself, we designed subsystems that use their own drive, emotions, and memory.

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 (Ledoux, J., 1996). While previous research on sociable robots either ignores the emotional memory or maintains the emotional memory implicitly in high-level interaction, we need to establish explicit memories of emotions, events, and concepts. Therefore we have 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.

Since 1999, AIM lab directed by Hyun S. yang in KAIST has focused on building new humanoid robots with a self-contained physical body, perception to a degree which allows the robot to be autonomous, and social interaction capabilities of an actual human symbiotic robot. This study was built on previous researches about the social interactions and the developments of the first generation humanoid robots, AMI, AMIET coupled with a human-size biped humanoid robot, AMIO in the AIM Lab (Yong-Ho Seo, Hyun S. Yang et al., 2003, 2004, 2006).

The proposed model was used to enhance the interaction between human and humanoid robot. Accordingly, we designed and implemented the methods which are mentioned above

and successfully verified the feasibility through the demonstrations of human and humanoid robot interactions with AIM Lab's humanoids.

2. Previous Sociable Robots and Affective Communication Model

In recent times, the concept of emotion has increasingly been used in interface and robot design. Moreover, numerous studies have been performed to integrate emotions into products including electronic games, toys, and software agents (C. Bartneck et al., 2001). Furthermore, Affective social interaction between human and robot is hot issues in robotics research area currently.

Many researchers in robotics have been exploring this issues such as a sociable robot, 'Kismet' which conveys intention through facial expression and engages in infant like interaction with a human caregiver (Breazeal, C., 1999). An AIBO, an entertainment robot, behaves like a friendly and life like dog which interact with human by touch and sound (Arkin and Fujita et al., 2003). Mel, a conversational robot, introduced concepts that a robot leads the interactions by explaining its knowledge (Sidner, C.L. et al., 2005). Cat Robot was developed to investigate the emotional behavior of physical interaction between a cat and a human (Shibata, T. at al., 2000).

Tosa & Nakastu have researched on emotion recognition system through speech and its applications. Their initial work, MUSE and MIC recognized human emotions involved in speech and expressed emotions through computer graphics. Since then, they developed some application systems utilizing their initial concept (Nakastu, Tosa et al., 1996,1999).

To perform high-level tasks while contending with various situations in actual, uncertain environments, a robot needs various abilities such as the ability to detect human faces and objects using a camera in addition to the speech processing, the ability to sense an obstacle using several sensors, and the capability of manipulation and bipedal navigation.

In addition, it is necessary to integrate these software functions efficiently and reliably. Therefore, to operate the robot, control architecture or framework was usually planned based on behavior architecture. The affective communication model which we designed is an unified control architecture to perform complex tasks successfully using a set of coordinated behaviors. A high-level task is driven from a motivation system. The motivation system activates set of coordinated behaviors. Finally, an expression system activates multimodal human interfaces such as a voice, gestures and 3D facial expressions.

3. Framework of Affective Communication Model

Motivated the human brain structure discovered by cognitive scientist (Ledoux, J., 1996), we have designed the framework for our sociable humanoid robots. 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 (Yoon, S.Y., Burke, R.C. et al., 2000). The similar framework was also applied to the software architecture of Kismet (Breazeal, C., 1999). 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 previous systems.

The first is a memory system. We added a memory system to the referred framework. The memory system enables a robot to represent, and reflect upon, itself and its human partners. It also enhances a robot's social skills and fosters communication with humans. To enable

affective interaction between a robot and humans, we enabled a robot to preserve its emotional memory of users and topics.

The second is a dialogue based human robot 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 a robot'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 (Yong-Ho Seo, Hyun S. Yang et al., 2003, 2004).

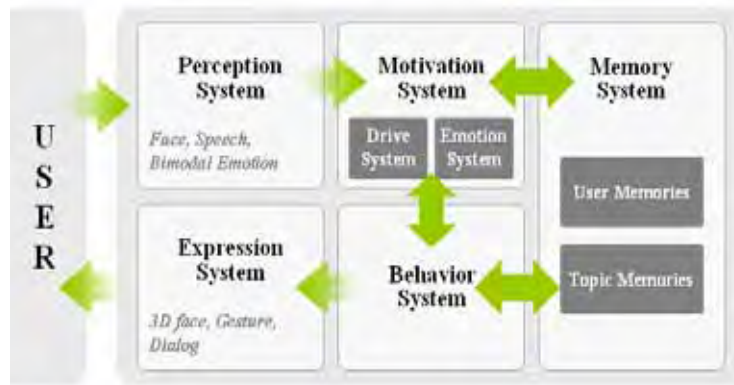


Figure 1. Framework of Affective communication model

4. Perception System

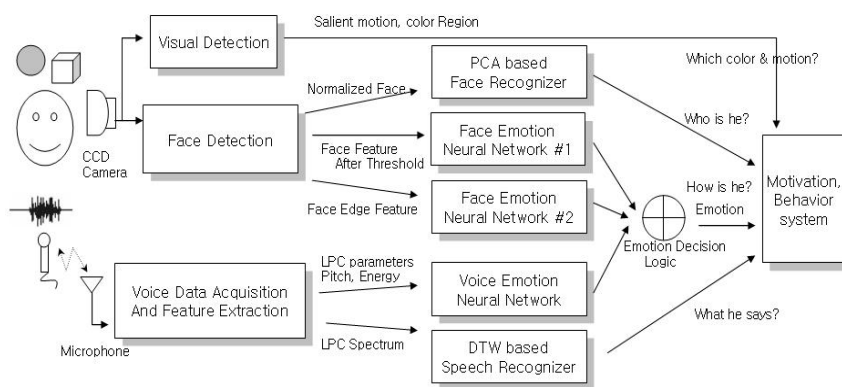


Figure 2. Structure of Perception system

The perception system comprises face detection, face recognition, speech recognition, emotion recognition, and visual attention detection that enables a robot to detect objects and color. Accordingly, through the perception system, the robot can learn basic knowledge such as the identity of the human user, the user's emotional state, and what the user is saying and doing. The overall process of the perception system is shown Fig. 2.

4.1 Face Detection and Recognition

The face detection system finds human faces in an image from CCD cameras. For robust and efficient face detection, the face detection system used a bottom-up, feature-based approach. The system searches the image for a set of facial features such as color and shape, and groups them 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 facial image is sent to the face recognizer and to the emotion recognizer. The face recognizer determines the user's identity from the face database. To implement the system, we used an unsupervised PCA-based face classifier commonly used in face recognition.

4.2 Bimodal Emotion Recognition

This section describes each emotion recognition and then bimodal emotion recognition and discusses the emotion recognition performance. We estimated emotion through facial expression and speech, and then integrate them to enable bimodal emotion recognition. Emotion recognition plays an important role in the perception system because our system enables a robot to recognize human partner's emotional status, and behave appropriately considering the recognized emotion status.

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 (Ekman, P. et al., 1978). The first feature is a facial image of lips, brow and forehead. After applying histogram equalization and the threshold of the standard distribution of bright of normalized face image, we extracted the parts of lips, brow and forehead from the entire image.

The second feature is an edge image of lips, brow and forehead. After applying histogram equalization, we extracted the edges around the regions of the lips, brow and forehead.

Each of the extracted features is then trained using two neural networks for five emotions (neutral, happy, sad, angry, and surprise). These emotions are chosen among the basic six emotions because people generally have difficulty in making facial expression especially for the emotions, fear and disgust, which show lower recognition rate than other four emotions. Each neural network is composed of 1610 input nodes, 6 hidden nodes and 5 output nodes. The 1610 input nodes receive 1610 pixels of the input image and the output nodes represent 5 emotions: neutral, happy, sad, angry, and surprise. The number of hidden nodes is determined by experiment.

Finally, the decision logic determines the final emotion result from two neural network results. The face emotion decision logic utilizes the weighted sum of two neural network results and voting method of result transitions over time domain.

The overall process of the emotion recognition through facial expression is shown in Fig. 3.

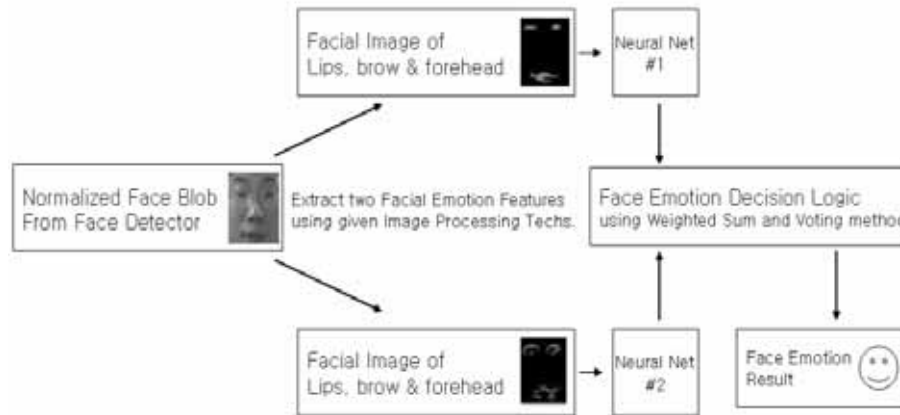


Figure 3. Process of Emotion Recognition through Facial Expression

The followings are about emotion recognition through speech. For emotion recognition through speech, we adopted a recognition method similar to the one used in the life-like communication agents MUSE and MIC (Nakatsu, Tosa et al., 1996, 1999). The speech features influenced by emotions are speech-influencing pitch, energy, timing, and articulation (Cahn, J., 1990). For feature extraction, we extracted phonetic and prosodic features in accordance with the state of the art research (Nakatsu, Tosa et al., 1999). Neural network is used to train each feature vector and recognize the emotions.

Two kinds of features are used in emotion recognition. One is a phonetic feature and the other is a prosodic feature. LPC (linear predictive coding) parameters, which are typical speech feature parameters often used for speech recognition, are adopted for the phonetic feature. The prosodic feature, on the other hand, consists of two factors: Speech power and pitch structure. Speech power and pitch parameters can be obtained in the LPC analysis. In addition, a delta LPC parameter is adopted, which is calculated from LPC parameters and expresses a time variable feature of the speech spectrum, since this parameter corresponds to a temporal structure.

The speech feature calculation is carried out in the following way. Analog speech is first transformed into digital speech by passing it through a sound card with an 8 KHz sampling rate and 16 bits accuracy. The digitized speech is then arranged into a series of frames, where each frame is a set of 256 consecutive sampled data points. LPC analysis is carried out in real time and the following feature parameters are obtained for each of these frames.

Thus, for the t -th frame, the obtained feature parameters can be expressed by

$$F_t = (P_t, p_t, d_t, c_{1t}, c_{2t}, \dots, c_{12t}) \quad (1)$$

(where speech power is P ; pitch is p ; Delta LPC parameter is d ; LPC parameters are $c_{1t}, c_{2t}, \dots, c_{12t}$.)

The sequence of this feature vector is fed into the speech period extraction stage.

For extraction of speech period, we used the information of speech power. Speech power is compared with a predetermined threshold value; if the input speech power exceeds this threshold value for a few consecutive frames, the speech is decided to be uttered. After the beginning of the speech period, the input speech power is also compared with the threshold

value; if the speech power is continuously below the threshold for another few consecutive frames, the speech is decided not to exist. The speech period is extracted from the whole data input through this process. Ten frames are extracted for the extracted speech period where each is situated periodically in the whole speech period and kept the same distance from adjacent frames. Let these ten frames be expressed as f_1, f_2, \dots, f_{10} . The feature parameters of these ten frames are collected and the output speech features are determined as a 150 (15x10) dimensional feature vector. This feature vector is expressed as

$$FV = (F_1, F_2, \dots, F_{10}) \quad (2)$$

(where F_i is a vector of the fifteen feature parameters corresponding to frame f_i .)

This feature vector (FV) is then used as input for the emotion recognition stage.

For emotion recognition, each feature vector was trained using a neural network for five emotions, neutral, happy, sad, and angry. The reason the four emotions was taken as classifiers is that these emotions have higher recognition rate than other emotions such as surprise, fear and disgust from the experimental results of previous systems to recognize emotion through speech.

Accordingly, the neural network was composed of 150 (15x10) input nodes corresponding to the dimension of speech features, 4 hidden node and 5 output nodes (neutral, sad, happy, angry and surprise). The neural network configuration is similar with the previous work of Nakatsu, R., Nicholson, J. and Tosa, N. (Nakatsu, Tosa et al., 1999), however internal learning weights of neural network and normalization algorithm of speech parameters are somewhat different.

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:

$$R_{final} = (R_{face}W_f + R_{speech}W_s) + R_{final-1} - \delta t \quad (3)$$

(where 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. δ is a decay term that restores the emotional status to neutral.)

4.3 Performance Evaluation of Bimodal Emotion Recognition

The overall correctness of the implemented 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.

We found that the two modalities have complementary property for each emotion from the experimental results shown in Table 1 and Table 2, as well as the previous research of bimodal emotion recognition (Huang, T.S., et al., 1998). The complementary property shows that for happy emotion, facial expression has higher recognition rates than speech. For sad emotion, voice has higher recognition rates. For angry emotion, facial expression and speech have similar recognition performance. Therefore, we made the final decision logic to conduct weighted summation. Accordingly, we appropriately determined the weight valuables, W_f and W_s in the reference of the experimental results.

Actually, it is difficult to extract hidden emotional features from natural facial expression and speech. Therefore, we used intentionally exaggerated facial expression and speech for each emotion to achieve high recognition rate compared to the previous research experiments.

For testing facial-only emotion recognition, we conducted experiments with four people. For training, we used five images per each emotion of each person. We set aside one from each category as test data, use the rest of samples as training data. The recognition result is shown in Table 1. Facial expression-only emotion recognition yield performance of 76.5% and 77.1% for the two neural networks. Therefore, we conducted weighted-summation to select the best result for each emotion from two neural networks and then achieved higher recognition rate of 79.5%.

Facial Expression - Neural Net. #1		Facial Expression - Neural Net. #2	
Happy	85.00 %	Happy	84.00 %
Sad	77.50 %	Sad	80.00 %
Neutral	77.50 %	Neutral	65.00 %
Angry	65.00 %	Angry	81.50 %
Surprise	77.50 %	Surprise	75.00 %
Total	76.5 %	Total	77.1 %

Table 1. Performance of Emotion from Facial Expression

In emotion recognition through facial expression, there is a little variation between people according to the Ekman's facial expression features [35]. On the other hand, there is a big difference in emotion recognition through speech because people have distinct and different voice. Especially, the speech features of men and women are largely different. Accordingly, we divided experiments into the two groups of men and women. In addition, we selected 4 emotions except surprise, because it is hard to recognize surprise from speech sentences. For four people (two men and two women), we trained 15 sentences frequently used in communication with the robot. The testers repeated one sentence for each emotion five times. We set aside one from each category as test data, used the rest of samples as training data. The average recognition rate of men and women is shown in Table 2.

Speech Expression - NN for Men		Speech Expression - NN for Women	
Happy	72.00 %	Happy	77.50 %
Sad	82.50 %	Sad	85.50 %
Neutral	75.50 %	Neutral	70.00 %
Angry	84.00 %	Angry	75.00 %
Total	78.5 %	Total	77 %

Table 2. Performance of Emotion from Speech Expression

The bimodal emotion system integrated facial and speech systems with one decision logic. We evaluated the bimodal system for four people in real-time environment with varying scales and orientations using a variety of complex backgrounds.

The participants were asked to make emotional facial expressions while speaking out the sentence emotionally for each emotion at five times during a specified period to time. The overall bimodal emotion system yielded approximately 80 % for each of four testers. It achieved higher performance results than facial-only and speech-only by resolving some confusion. The higher result of this emotion recognition system compared to the other systems is caused by the limited number of users. Therefore, if the more users are

participated in this recognition system, the lower recognition result is expected. It's the limitation of these emotion recognition systems.

5. Motivation System

The motivation system sets up the robot's nature by defining its "needs" and influencing how and when it acts to satisfy them. The nature of the robot is to affectively communicate with humans and ultimately to ingratiate itself with them. The motivation system consists of two related subsystems, one that implements drives and a second that implements emotions. Each subsystem serves as a regulatory function for the robot to maintain its "well-being"

5.1 Drive System

The motivation system defines the robot's nature by defining its "needs" and influencing how and when it acts to satisfy them. The nature of the proposed humanoid robot is to socially interact with humans and ultimately to ingratiate itself with them. The motivation system consists of two related subsystems, one that implements drives and a second that implements emotions. Each subsystem serves as a regulatory function for the robot to maintain its "well-being"

In our previous research, three basic drives were defined for a robot's affective communication with humans (Yong-Ho Seo, Hyun S. Yang et al., 2004). In the new drive system for a humanoid robot operating and engaging interactions with human, four basic drives were defined for the robot's objectives as they related to social interaction with a human: a drive to obey a human's commands; a drive to interact with a human; a drive to ingratiate itself with humans and a drive to maintain its own well-being.

The first drive motivates a robot to perform a number of predefined services according to a human's commands. The second drive activates the robot to approach and greet humans. The third drive prompts the robot to try to improve a human's feelings. When the robot interacts with humans, it tries to ingratiate itself while considering the human's emotional state. The fourth drive is related to robot's maintenance of its own well-being. When the robot's sensors tell it that extreme anger or sadness is appropriate, or when its battery is too low, it stops interacting with humans

5.2 Emotion System

Emotions are significant in human behavior, communication and interaction (Armon-Jones, C., 1985). A synthesized emotion influences the behavior system and the drive system as a control mechanism. To enable a robot to synthesize emotions, we used a model that comprises the three dimensions of emotion (Schlossberg, H., 1954). 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 a robot is always openly involved in interactions. Therefore, we only consider valence and arousal, implying that only three emotions are possible for our robots: happiness, sadness, and anger.

The arousal factor (Arousal about current user) is determined by factors such as whether a robot finds the human, and whether the human responds. Low arousal increases the emotion of sadness.

The valence factor (Response about current user) is determined by whether the human responds appropriately to robot'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. The robot's emotional status is computed by the following equation.

$$\begin{aligned} \text{If } t = 0, E_i(t) &= M_i \quad (t = 0 \text{ when new face appears}) \\ \text{If } t \neq 0, E_i(t) &= A_i(t) + E_i(t-1) + \sum D_i(t) + M_i - \delta t. \end{aligned} \quad (4)$$

Where $E_i(t)$ is the robot's emotional status, t is time(when new face appears), $i = \{\text{joy, sorrow, anger}\}$. $A_i(t)$ is the emotional status calculated by the mapping function of $[A, V, S]$ from the current activated behavior. D_i is the emotional status defined by the activation and the intensity of unsatisfied drives in the drive system. M_i is the emotional status of the human recorded in the memory system. Finally, δt is a decay term that eventually restores the emotional status to neutral.

6. Memory System

Topic memories contain conversational sentences that a robot 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 the behavior system confirms that the current 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 robot's current emotion.

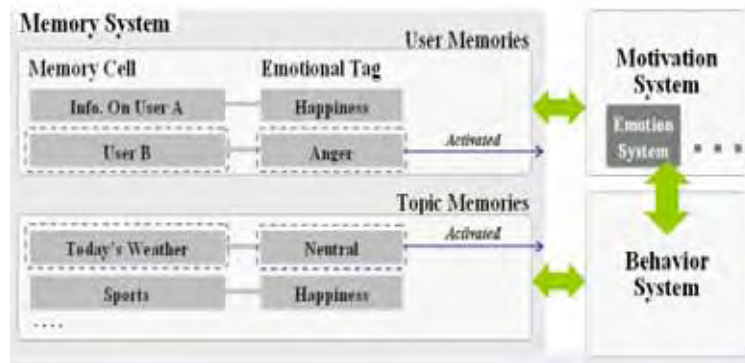


Figure 4. Activation of Memory cells in the Memory System

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:

$$\begin{aligned} \text{If } COMM = 0, T_i(t) &= W_{mt} \sum E_k(t) ET_i(t) \\ \text{If } COMM = i, T_i(t) &= 1 \end{aligned} \quad (5)$$

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 robot's current emotion and the emotion of the memory of the topic. Finally, W_{mt} is a weight factor. The activation of the memory system is shown in following Fig. 4.

7. Behavior and Expression System

We designed the structure of the behavior system that has three levels, which address the three drives of the motivation system as mentioned above. 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. The Fig. 5 shows the hierarchy of the behavior system and its details.

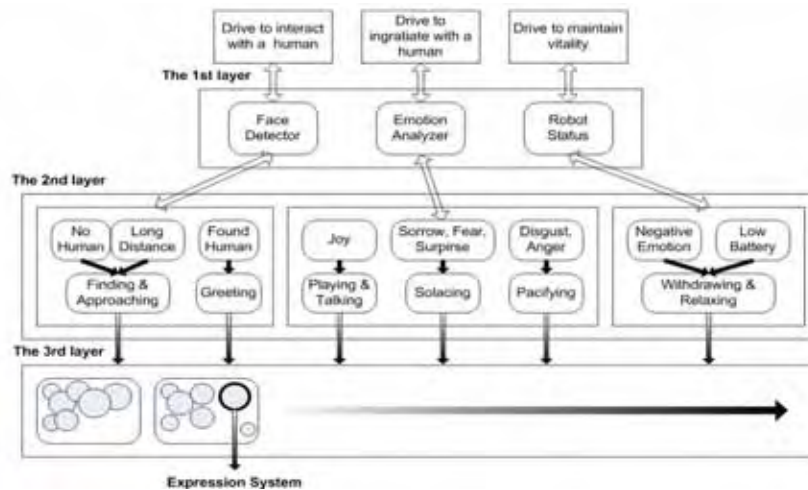


Figure 5. Hierarchy of the Behavior System

The expression system is the intermediate interface between the behavior system and robot hardware. The expression system comprises three subsystems: a dialogue expression system, a 3D facial emotion expression system and a gesture expression system.

The expression system plays two important functions. The first function is to execute the behavior received from the behavior system. Each type of behavior consists of a dialogue between the robot and the human. Sometimes the robot uses interesting gestures to control the dialogue's flow and to foster interaction with the human. The second function is to express robot's emotion. The robot expresses its own emotions through facial expressions but it sometimes uses gestures to convey its intentions and emotions.

7.1 Dialogue Expression

Dialogue is a joint process of communication sharing of information (data, symbols, context) between two or more parties. In addition, humans employ a variety of paralinguistic social cues (facial displays, gestures, etc.) to regulate the flow of dialogue (M. Lansdale, T. Ormerod, 1994). We consider there to be three primary types of dialogue: low level (prelinguistic), non verbal, and verbal language. Among them, the robot communicates with a human through daily verbal language with appropriate gestures.

However, it is difficult to enable a robot to engage in natural dialogue with a human because of the limitation of the current technique on speech recognition, natural language processing, etc. Accordingly, we predefined dialogue flow and topics. In order to make natural dialog possible in the limit that the robot could recognize only the limited number of speech, we constructed dialog as follows: First, the robot lead actively dialog by asking user's intention ahead of him to avoid the possibility that the robot cannot understand human speech. Second, the robot answers the most frequently used responses when they cannot understand to avoid unnatural dialog.

The dialogue expressions comprise the most commonly used speeches according to the selected behavior of Finding & Approaching, Greeting, Talking, Playing, Consoling, Pacifying, Withdrawing & Resting behavior group. In Finding & Approaching, the robot mainly finds a human by calling a human. In greeting, the robot says hello to a human and asks his name, etc. In Talking, dialog consists of various common topics such as hobby, weather, movies, etc. In playing, the robot plays with a human through a kind of jokes, OX quiz and Nonsense Quiz. In Consoling and Pacifying, the robot asks what the human is angry about and then makes a joke to console him or give pleasure to him. Further, the robot asks his worries by recognizing his sad emotion and listens to his sayings as his friend and counselor. The followings are the examples of the conversations in human robot interaction using the proposed model.

Sample conversation between AMI and users

Conversation 1

(GREETING)

AMI: Is there anyone who wants to talk to me?

User 1: Hello, AMI.

User 1 approaches. AMI finds a human face.

AMI: I heard someone calling me.

AMI: Ah, you are there. You're (User 1's name).

AMI: Let's shake hands.

User 1: OK.

User 1 and AMI shake hands.

AMI: How are you today?

User 1: Not very good. I had a pretty bad day.

AMI: I'm sorry to hear that.

(TALKING)

AMI: Is it cold today?

User 1: Yes.

AMI: I think so. I like warm and sunny days. How about you?

User 1: Well, I also prefer warm weather.

AMI: I thought so.

(CONSONLING)

AMI: Anyway, you look somewhat sad. Is there anything wrong with you?

User 1: Yes.

AMI: Yeah, you look pretty unhappy. Can you tell me what happened to you today?

User 1: OK.

AMI: Tell me then.

User 1: It was a terrible day for me. I had a lot of homework to submit and I also had an exam today. I didn't have enough time to study for the exam because of my homework. So my exam was terrible, and that's why I'm not feeling good now.

AMI: I really feel sorry about that. Why don't we listen to some music?

User 1: Good idea.

AMI plays some music.

... (The conversation continues until the user says, "AMI, let's stop talking.")

Conversation 2

(GREETING)

AMI: Hello? Is anyone there?

User 2: Hi.

User 2 starts speaking.

AMI: Ah. I heard a voice. Where are you?

User 2: Turn left.

AMI: I found you. Nice to meet you. Can you tell me your name?

User 2: (User 2's name).

AMI: You're (User 2's name). Let's shake hands.

User 2: OK.

User 2 and AMI shake hands.

AMI: How are you today?

User 2: Fine.

AMI: I'm fine too.

(TALKING)

AMI: What did you have for lunch?

User 2: I had a hamburger, potato chips, and a salad.

AMI: Sounds nice. What's your favorite food?

User 2: Hmm. Nothing particular. I like everything.

AMI: Good.

(PLAYING)

AMI: You look fine today. Let's play a game.

User 2: OK.

AMI: Good. How about a common sense quiz. I'll give you some questions and you have to answer. Is that OK?

User 2: OK.

AMI: The first world cup was held in Greece.

User 2: Correct.

AMI: Sorry. Wrong. Next. The reason for VDT syndrome is too much exercise.

User 2: No!

AMI: You're right. Good job.

User 2: Had enough?

AMI: OK. Let's do it again later.

... (The conversation continues until the user says, "AMI, let's stop talking.")

7.2 Facial Expression

The 3D facial expression shows the robot's emotional status synthesized in the motivation system, as described in section 5. These expressions make up for the limitations of the robot's mechanical face which has difficulty in expressing its emotions. These facial emotion expressions were implemented using 3D graphics. Our 3D graphical face is displayed on the LCD screen which located on the robot's chest. We developed two different facial expression programs. One is more face like version and the other is more artificial and abstract version. The facial expressions in our 3D graphical faces and the dimension of emotions are shown as Fig. 6.

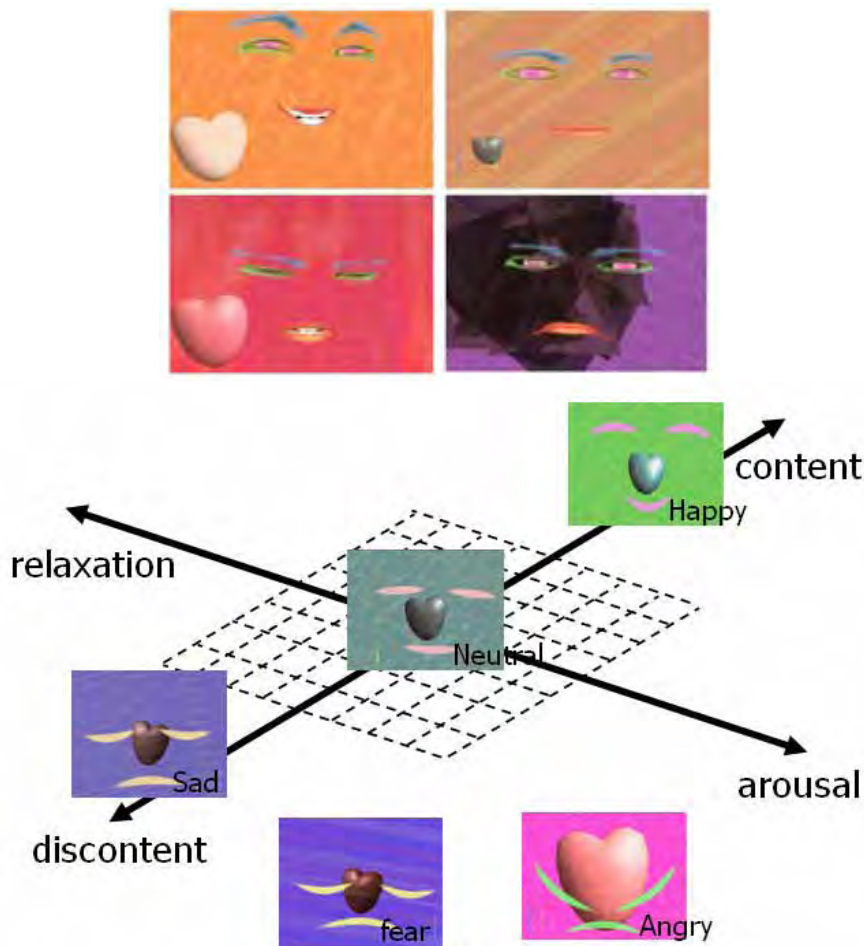


Figure 6. Graphical Facial Emotion Expressions and Dimension of Emotions

7.3 Emotional Gesture Expression

Gestures(or Motions) for our humanoids were generated to be human-like and friendly. Gestures are used to express its own emotions and to make interaction with humans more expressive. Therefore, expressions that would best attract the interest of humans were considered, and various interesting gestures were developed for our humanoids that would match the robot's dialogs and emotional statuses.

Humans tend to guess the emotional states of other people or some object from their body motions. Motions of a service robot are also important because they give strong impressions to a person. Most people think that robots act unnaturally and strangely. There are three types of functional disorders in communication methods between a human and a robot excluding speech. The details are in Table 3.

Limitations of conventional robots	Functional disorders
Motions to express internal state	<p>Problem</p> <p>Conventional robots can not express their internal state ex) out of battery, emergency</p> <p>Solution</p> <p>Motions can be used for expressing internal state of a robot ex) no movement - out of battery slow movement - ready fast movement - emergency</p>
Communication using sense of touch	<p>Problem</p> <p>No reactions when a robot is touched Ex) An accident can be occurred even though someone tries to stop the robot.</p> <p>Solution</p> <p>A robot can express its internal state using motions when it touched by others Ex) When a person punishes a robot for its fault by hitting it, it trembles.</p>
Eye Contact	<p>Problem</p> <p>A robot which has no eyes looks dangerous ex) Humans usually feel that robots with no eyes are dangerous</p> <p>Solution</p> <p>A robot can look at a person of interest with sense of vision. ex) When a robot is listening to its master, it looks at his/her eyes.</p>

Table 3. Limitations of conventional robots' interaction

We have to improve above functions of a robot to express its internal emotional state. As we can see Table 3, these functions can be implemented by using the channels of touch and vision. We focused the channel of vision perception-especially motion cues, so we studied about how to express emotions of a robot using motions such as postures, gestures and dances.

To generate emotional motions of a service robot, making an algorithm which can convert emotion to motions and describing motions quantitatively are necessary. We defined some parameters to generate emotional motions. These parameters are like in Table 4. We defined the parameters for the body part and the parameters for the two arms independently, so we can apply these parameters to a robot without considering whether it has two arms or not. Posture control and velocity control are very important to express emotional state using activities. These parameters are not absolute values, but relative values.

	Parameter	Joy	Sad	Anger	Disgust	Surprise
Body	Velocity	Fast	Slow	Fast	Slow	Slow
	Acceleration	Small	-	Large	Small	Large
	Direction	Possible turns	-	Forward / Backward	Backward	Backward / Stop
Arms	Position	Up	Down	Center	Center	Up
	Velocity	Fast	Slow	Fast	Normal	Fast
	Velocity change	Small	-	Large	Small	Small
	Shape	Arc	Line	Perpendicular	Perpendicular	Perpendicular
	Symmetry	Symmetrical	-	Unsymmetrical	-	-

Table 4. Parameters for emotional motions

To generate emotional gestures, we used the concept of Laban Movement Analysis, which is used for describing body movements (Toru Nakata, Taketoshi Mori, et al., 2002).

There are various parameters which are related to produce emotional motion generation. These parameters are related to generating natural emotional motions. To make natural motions of a robot, these parameters are used for expressing the intensity of the emotional state. According to the intensity of emotion, the number of parameters is changed to generate emotional motions. The higher intensity of an emotion is going to be expressed in a motion, the more parameters are going to be used for generating that motion.

We described the details of the parameters for emotional motions and we developed the emotional motions generating method. We defined 8 parameters and we can express 6

emotions by adjusting these parameters. The emotions we can express are joy, sad, neutral, surprise and disgust. We can make emotional motions with 5 levels of intensity by adjusting parameters in Table 2. We developed a simulator program to preview the generated motions before applying to the robot platform. The simulator is shown in Figure 7.



Figure 7. Simulator for emotional motion generation

We can preview a new generated motion using this simulator, so we can prevent some problems which can be occurred when we try to apply that motion to the robot. In this simulator, we can produce 6 emotional motions. Each emotional motion has 5 levels corresponding to the intensity of the emotion. Some examples of these emotional motions of our humanoid robots are shown in Fig. 8, Fig. 9, respectively.

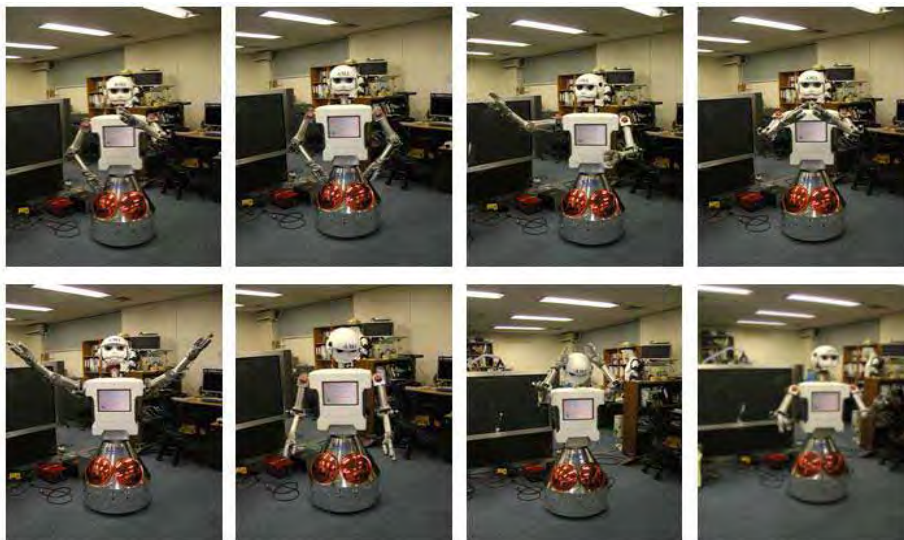


Figure 8. Gesture expressions of AMI

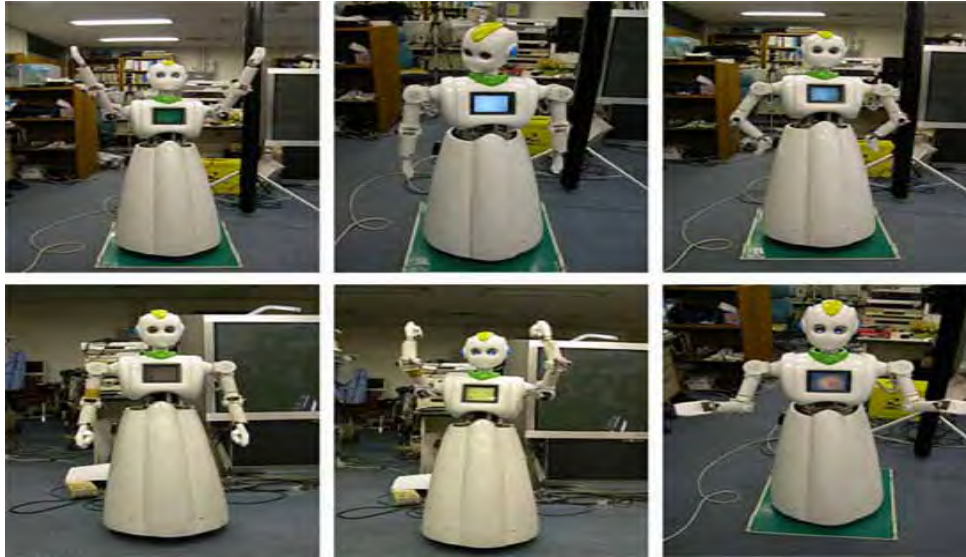


Figure 9. Gesture expressions of AMIET- joy, sad, anger, neutral, surprise and disgust in sequence

8. AIM Lab's Humanoid Robots

This section summarizes the study on design and development of the humanoid robots of AIM Lab to realize the enhanced interaction with humans. Especially, we have been focusing on building a new robot with the self-contained physical body, the intelligence which make the robot be autonomous, and the emotional communication capability toward a human symbiotic robot in AIM Lab, since 1999.

So far, the members of AIM Lab have developed autonomous robots called AMI and AMIET which have two wheeled mobile platform, anthropomorphic head, arms and hands. And also we have been developing software system which performs intelligent tasks using a unified control architecture based on behavior architecture and emotional communication interfaces. Humanoid robots, AMI, AMIET were released to the public in 2001, 2002 respectively.

AMIO is the biped humanoid robot which was developed recently. The developed robot consists of a self-contained body, head, two arms, with a two legged (biped) mechanism. Its control hardware includes vision and speech capabilities and various control boards such as motion controllers, with a signal processing board for several types of sensors. Using the developed robot, biped locomotion study and social interaction research were concurrently carried out.

An anthropomorphic shape is appropriate for a human-robot interaction. Also it is very important that the robot has the stable mobility in dynamically changing and uncertain environment. Therefore, we decided the design of our robot as the mobile manipulation robot system which has upper torso, head, two arm, hand, and vehicle. In the first mechanical design stage of AMI, we consider the following factors for our robot to satisfy the requirements of these human symbiotic situations.

We considered the height of the robot firstly. The height of the robot should be appropriate to interact with human. Secondly, Manipulation capability and motion range of robot arm should be similar to human. Thirdly, Reliable mobility to move around household while ensuring human's safety is required. Fourthly, Information Screen to show helpful information to user through the network environments and to check the current status of robot and to show emotional expression of robot itself is required. In the consideration of these points, we tried to make our robot be more natural and intimate to human. And then we built the robots, AMI and AMIET as following Fig. 10.

The designed robot, AMI has a mouth to open and close in the case of speaking and making expressions, two eyes with CCD camera, and speaker unit to make sounds in his face. And his neck is equipped with two motors to track the moving target and implement active vision capabilities.

The developed torso part supports the robot's head and two arms, and includes the main computer system, arm and head controllers, and motor driving units. And also, LCD screen is attached to his breast to check the internal status of robot and to recover the limitations of mechanical face which has difficult in making emotional expressions.

We designed two symmetric arms which have five degrees of freedom each. Hand has six degrees of freedom and three fingers. Each finger has two motors. At the end of fingers, FSR(Force Sensing Register) sensors are located to sense the force in grasping object. The total length of AMI's arm with hand is 750[mm].

In case of AMIET, it is designed to make human-like motions with its two arms; each arm has 6 degrees of freedom (DOF), so it can imitate the motion of a human arm. Additionally, AMIET has a waist with 2 DOF to perform rotating and bending motions. Thus, AMIET can perform many human-like acts.

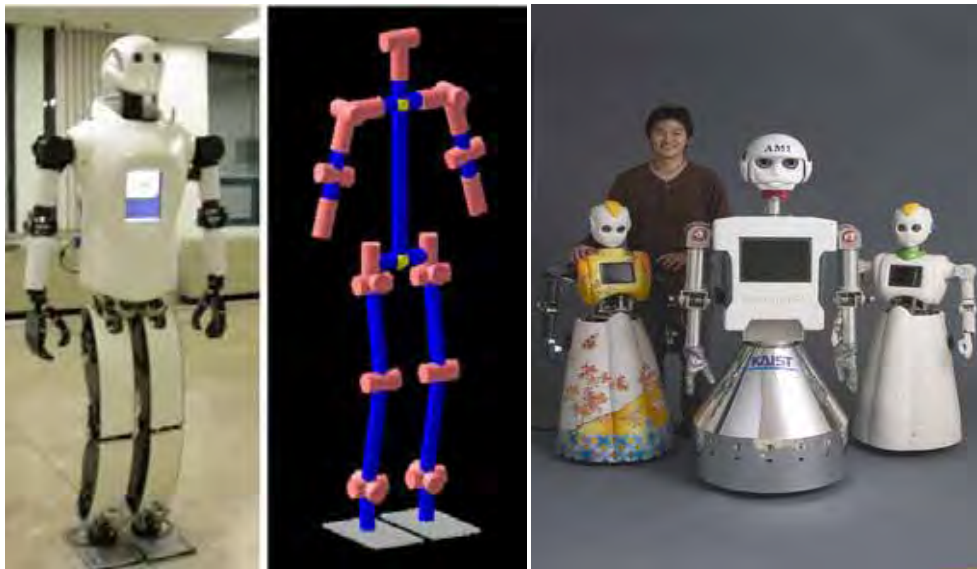


Figure 10. AIM Lab's Biped Humanoid Robots, AMIO, AMI and two AMIETs

AMI is 1550 mm tall. The breadth of the shoulder is 650 mm and the total weight is about 100 kg. Figure 13 shows the shape and DOFs of the assembled robot. AMIET has child-like height. Ami is 160 cm tall and AMIET is 130 cm tall. Differently with AMI, AMIET is designed by the concept of a child and her appearance is considered before when she is developed. So AMET could be felt friendly for humans.

A newly developed biped humanoid robot named AMI was designed and manufactured based on the dimensions of the human body. The lower part of the robot has two legs, which have 3, 1, and 2 degrees of freedom at the pelvis, knees, and ankles, respectively. This allows the robot to lift and spread its legs, and to bend forward at the waist. This structure, which was verified by previous research to be simple and stable for biped-walking robots, makes it possible for the robot to walk as humans walk. The shape and D.O.F arrangement of the developed robot, AMIO are shown in Fig. 10.

9. Experimental Results

To test the performance of the affective communication model, we conducted several experiments with humanoid robots, AMI, AMIET and AMIO. We confirmed that each subsystem satisfies its objectives. From our evaluation, we drew the graph in Fig. 12, which shows the subsystem's flow during a sample inter-action. The graph also shows the behavior system (finding, greeting, consoling and so on), the motivation system (robot's drives and emotions), and the perception system (the user's emotional status)

To evaluate the proposed communication model with 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 5 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 the robot to synthesize more natural emotions and to reduce redundancy in conversational topics. Fig. 13 and Fig. 14 show the cases of the human and humanoid robot interaction experiments using the proposed affective communication model.

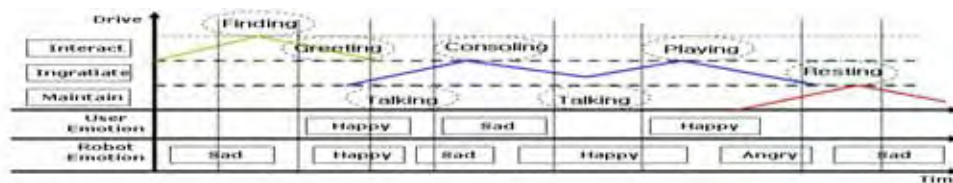


Figure 12 Work flow of the system

		without emotion, without memory	with emotion, without memory	with emotion, with memory
Natural		54%	70%	76%
Unnatural	Emotion mismatch	19%	10%	5%
	Redundancy	14%	6%	3%
	Other errors	13%	14%	16%

Table 5. Experimental results of Affective interactions with a robot



Figure 13 AMI's Affective Dialogue Interactions with Mr. Ryu in TV program



Figure 14 AMIO and AMIET shake hands with human in Affective Interaction Experiment

10. Conclusion

This chapter presented the affective communication model for humanoids that is designed to lead human robot interactions by recognizing human emotional status and expressing its emotion through multimodal emotion channels like a human, and behaves appropriately in response to human emotions.

We designed and implemented an affective human-robot communication model for a humanoid robot, which makes a robot communicate with a human through dialogue. Through this proposed model, a humanoid robot can communicate with humans by preserving emotional memories of users and topics, and it naturally engages in dialogue with humans.

With explicit emotional memories on users and topics, in the proposed system, we successfully improved the affective interaction between humans and robots. Previous sociable robots either ignored emotional memories or maintained them implicitly. Our research proves that explicit emotional memory can help high-level affective dialogue interactions.

In several experiments, the robots chose an appropriate conversation topic and behaved appropriately in response to human emotions. They could ask what the human is angry about and then make a joke to console him or give pleasure to him. Therefore, this robot is able to help human mentally and emotionally as a robot therapy function. The human

partner perceives the robot to be more human-like and friendly, thus enhancing the interaction between the robot and human.

In the future, we plan to extend the robot's memory system to contain more various memories, such as visual objects or high level concepts. The robot's current memory cells are limited to conversational topics and users. Our future system will be capable of memorizing information on visual inputs and word segments, and connections between them.

To interact socially with a human, we are going to concentrate on building a real humanoid robot in terms of thinking and feeling that can not only recognize and express emotions like a human, but also share emotional experience with humans while the robot is talking to users on many kinds of interesting and meaningful scenarios supported and updated dynamically from outside database systems such as worldwide web and network based contents server.

Acknowledgement

This research was partially supported by the Korea Ministry of Commerce, Industry and Energy(MOCIE) through the Brain Science Research Project and Health-Care Robot Project and by the Korea Ministry of Science and Technology(MOST) through AITRC program.

11. References

- Armon-Jones, C. (1985): The social functions of emotions. R. Harre (ed.), *The Social Construction of Emotions*, Basil Blackwell, Oxford.
- Arkin, R.C., Fujita, M., Takagi, T., Hasegawa, R. (2003): An Ethological and Emotional Basis for Human-Robot Interaction. *Robotics and Autonomous Systems*, 42.
- Breazeal, C. and Scassellati, B. (1999), A context-dependent attention system for a social robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pp.1146-1151.
- C. Bartneck, M. Okada (2001), Robotic user interfaces, *Proceedings of the Human and Computer Conference*, 2001.
- Cahn, J. (1990), Generating expression in synthesized speech, *Master's Thesis*, MIT Media Lab.
- Ekman, P., Friesen, W.V. (1978): *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, Palo Alto, CA.
- Huang, T.S., Chen, L.S., and Toa, H. (1998), Bimodal Emotion Recognition by Man and Machine. *ATR Workshop on Virtual Communication Environments*. 1998
- Hyun S. Yang, Yong-Ho Seo, Yeong-Nam Chae, Il-Woong Jeong, Won-Hyung Kang and Ju-Ho Lee (2006), Design and Development of Biped Humanoid Robot, AMI2, for Social Interaction with Humans, *proceedings of IEEE-RAS HUMANOIDS 2006*
- Jung, H., Yongho Seo, Ryoo, M.S., Yang, H.S. (2004): Affective communication system with multimodality for the humanoid robot AMI. *proceedings of IEEE-RAS HUMANOIDS 2004*
- Ledoux, J. (1996): *The Emotional brain: the mysterious under pinning of emotional life*. New York: Simon & Schuster
- M. Lansdale, T. Ormerod (1994), *Understanding Interfaces*, Academic Press, New York
- Naoko Tosa and Ryohei Nakatsu (1996), *Life-like Communication Agent - Emotion Sensing Character "MIC" & Feeling Session Character "MUSE"*, ICMCS, 1996

- Nakatsu, R., Nicholson, J. and Tosa, N. (1999), Emotion Recognition and Its Application to Computer Agents with Spontaneous Interactive Capabilities, *Proc. of the IEEE Int. Workshop on Multimedia Signal Processing*, pp. 439-444, 1999
- Ledoux, J. (1996): *The Emotional brain: the mysterious under pinning of emotional life*. New York: Simon & Schuster.
- Schlossberg, H. (1954): Three dimensions of emotion. *Psychology Review* 61
- Shibata, T. et al. (2000): Emergence of emotional behavior through physical interaction between human and artificial emotional creature. *ICRA (2000)* 2868-2873
- Sidner, C.L.; Lee, C.; Kidds, C.D.; Lesh, N.; Rich, C. (2005), Explorations in Engagement for Humans and Robots, *Artificial Intelligence*, May 2005
- Toru Nakata, Taketoshi Mori & Tomomasa Sato (2002), Analysis of Impression of Robot Bodily Expression, *Journal of Robotics and Mechatronics*, Vol.14, No.1, pp.27--36, 2002
- Yong-Ho Seo, Ho-Yeon Choi, Il-Woong Jeong, and Hyun S. Yang (2003), Design and Development of Humanoid Robot AMI for Emotional Communication and Intelligent Housework, *Proceedings of IEEE-RAS HUMANOIDS 2003*, pp.42.
- Yoon, S.Y., Burke, R.C., Blumberg, B.M., Schneider, G.E. (2000): Interactive Training for Synthetic Characters. *AAAI 2000*