PROCEEDINGS OF THE 2ND COE WORKSHOP ON HUAN ADAPTIVE MECHATRONICS (HAM) — INVITED SPEAKERS VERSION —

Date: Friday and Saturday, 4 - 5 March 2005
Place: Tokyo Denki University (Hatoyama Campus)
Hatoyama-machi, Hiki-gun
Saitama 350-0394, Japan

Organized by the Center of Excellence (COE) Project
Tokyo Denki University

The COE Project Office, Tokyo Denki University
Hatoyama-machi, Hiki-gun, Saitama 350-0394, Japan

http://www.ham.coe.dendai.ac.jp
E-mail: COE.office@ham.coe.dendai.ac.jp

Proceedings of the 2nd COE Workshop on Human Adaptive Mechatronics(HAM)
4-5 March 2005, Tokyo Denki University, Japan

A Human-Robot Interaction Framework for Home Service Robots

KangWoo Lee[*,***], Hyoung-Rock Kim[**], Dong-Soo Kwon[**,***], Wan Chul Yoon[*,***]

[*]Industrial Engineering, Korea Advanced Institute of Science and Technology
[**]Mechanical Engineering, Korea Advanced Institute of Science and Technology
[***]Human Robot Interaction Research Center, Korea Advanced Institute of Science and Technology
Guseong-dong, Yuseong-gu, Daejeon, Korea
kwonds@kaist.ac.kr

Abstract. This paper presents a human-robot interaction framework that outlines a general structure of future home service robots that are expected to assist humans in their home-based daily activities. We describe three interaction modules – multimodal, cognitive, and emotional interaction modules. Each module takes a different role in the process of human robot interaction. The multi-modal interaction is to make the interaction convenient for the human, the cognitive interaction is for cooperative sharing of tasks, and the emotional interaction is to maintain a close relationship. The general concept for the systematical software integration and the relationships among three modules is described.

Keywords. *Human-Robot Interaction (HRI), Multi-modal, Cognitive model, Emotional model, Service Robot*

I. INTRODUCTION

The ecological niche of robots is extending from industrial fields to our living and working places such as homes, offices, restaurants, etc. As it moves toward human's habitat, the interaction between human and robot is inevitably required, and becomes a key issue in robotics and its related research areas including psychology, artificial intelligence, ergonomics etc. The transfer of robot's ecological niche from industrial field to human's niches is accompanied not only with the change of a robot's task, but also with the change of interaction patterns between human and the robot. These changes call for more active challenges at least in three research areas – multi-modal, cognition, and emotional interaction.

Robotics researchers begin to change their views on robots from as a tool or device operated by a human user to as a partner or friend in the workplace or home. The work on the issues such as how to acquire socially sophisticated relationship, how to establish collaborative relationship, and how to exchange affective states has already been started. Some experimental trials have been carried out to answer those questions. For example, Cog and Kismet have been developed in order to study and realize a sensible bi-directional social relationship between human and robot based on psychological and biological studies [1, 2, 3, 4]. The Cero has been used to investigate the interactive capabilities under various cognitive factors such as user's procedural knowledge about a task, user's expectation etc [5, 6]. Various receptive and expressive communication channels have been adapted to realize natural interaction, and also to investigate the effectiveness of multimodalities in social exchange.

In this paper we introduce our research endeavor on human-robot interaction (HRI) at the Korea Advanced Institute of Science and Technology (KAIST). We have launched HRI research project as a part of Intelligent Robotics Development Program (IRDP). Researchers in various academic areas such as robotics, computer science, psychology, cognitive engineering etc. participate in the project and cooperate in order to develop a home service robot that is able to cooperate with its human companion, and assist his work.

In this paper, we first outline a general framework of our approach to HRI in Section II. The general framework explains why approaches of HRI are different from the approaches of artificial intelligence (AI), or human-computer interaction (HCI). The framework also focuses 3 interaction modules – multimodal, cognitive and emotional interaction module. These modules are described in detail in section III. Section IV describes how to integrate these three modules into a single system and how one of them interacts with another. Future research directions and plans are also presented in brief.

II. HRI FRAMEWORK

2.1 Approach HRI

HRI is somewhat different from AI, which directly or loosely models human intelligence. HRI involves modeling of the linkage that connects human and robot intelligence, not being limited to human intelligence alone. The following figure explains the relationship with AI in terms of modeling.

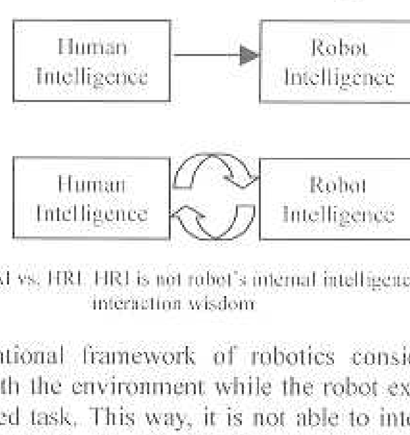

Figure 1. AI vs. HRI. HRI is not robot's internal intelligence, but interaction wisdom

The conventional framework of robotics considers the interaction with the environment while the robot executes a preprogrammed task. This way, it is not able to interpret of how HRI occurs in the context of a world domain. Our approach to HRI tries to establish the common grounds shared by human and robot. In principle, the mutual interaction implies there is a perceptual, cognitive, emotional, or social ground shared by the agents. The establishment of the common ground is a critical premise that allows us to interact with robots. For instance, the perceptual abilities of robots are on a strictly metric (or quantitative) level, whereas human perception is based on the qualitative categorization. So, the color 'red' is differently represented in two systems – RGB values in robot systems and coarse color representation in human systems. The implementation of both quantitative and qualitative perceptual ability builds up the common grounds that links human and robot perceptual experience.

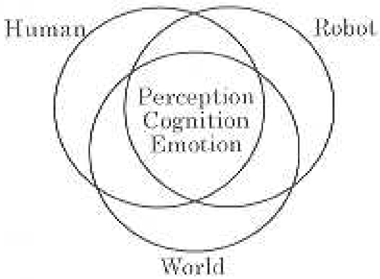

Figure 2. Shared grounds by human and robot

In addition, the HCI (Human Computer Interaction) framework is also not suitable for modeling the interactive relationship between human and robot [7] as computer is not autonomous and less interactive, whereas robot has more autonomy, cognitive or emotional abilities in order to realize bi-directional, implicit, and affective relationship with the human companion. Therefore, a different approach is obviously required to theorize what a relationship between human and robot might be, what knowledge should be shared, what knowledge about the other can be modeled etc.

2.2 Three Interaction Modules

The framework we propose is based on three interaction modules: multi-modal, cognitive interaction module, and emotional interaction module that are involved in different (but related) aspects of HRI. The division between multimodal interaction and the other two modules is based on the level of information processing stages. The multimodal interaction uses various communication channels to comprehend the intention or the emotional states of the human companion, and also to deliver robot's own response. The integration of multiple information obtained from different sensory modalities and maintenance of coherence between expressive channels are the core of the modules. Even though there is no clear distinction between cognition and emotion, the distinction between cognitive and emotional interaction in our approach lies on the orientation of interactivity. The interaction attributes in cognitive module are more task-oriented, whereas the interaction attributes in emotional module are more human oriented (For more details, see section III). Moreover, the interaction in cognitive module is for the cooperative relation with a user to achieve a goal of a task, whereas that in emotional module is for maintaining

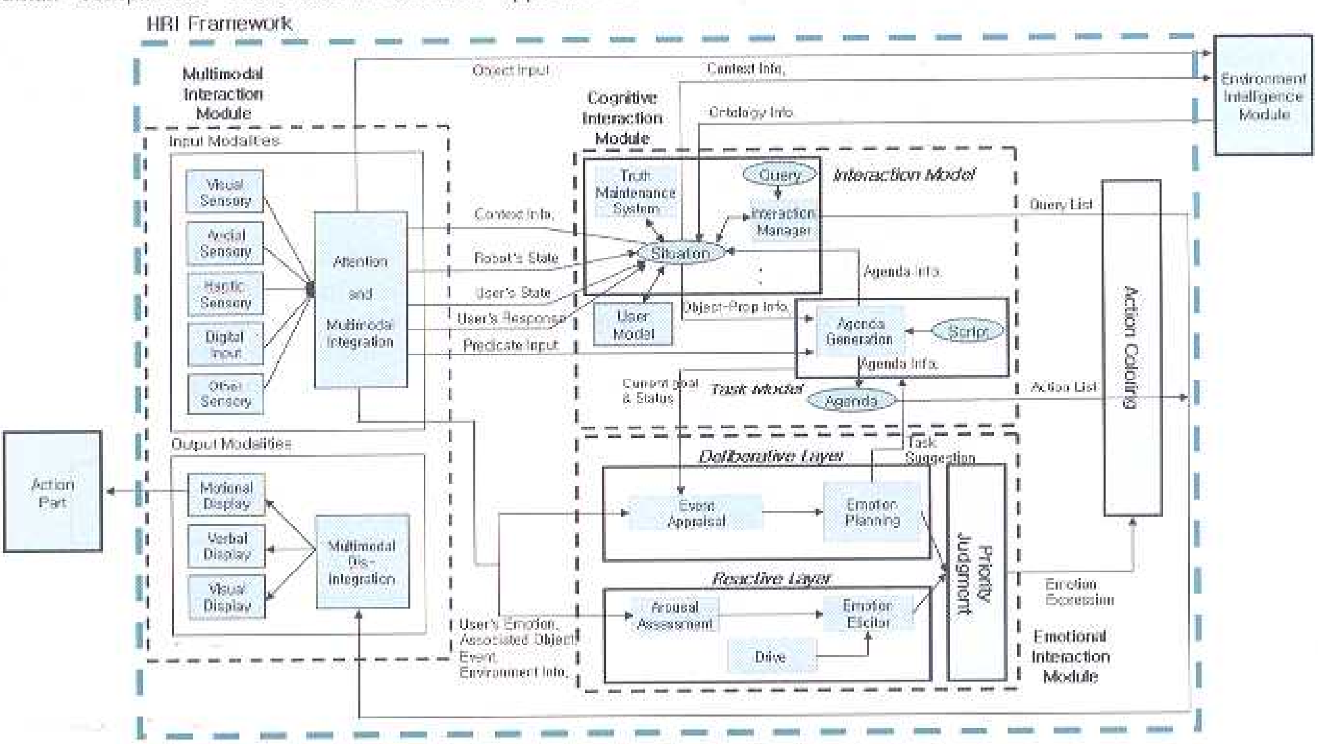

Figure 3. HRI Framework in the Robot. The thick and dashed box represents HRI framework that we present in this paper. The framework consists of 3 interaction modules – multimodal, cognitive and emotional interaction modules. The distinction between 3 modules lies on the level of information process stages and orientation of interactivity. The multimodal interaction module plays a role of interface to receive multimodal inputs and to generate or express actions. The interaction attributes in cognitive interaction module is more task-oriented, where those in emotional model are more human-oriented. For more details, see section III.

social relation with a human user. Even though there are on-going debates of whether personality or social intelligence need to be implemented in a service robot, these can be certainly beneficial to comprehend the emotional state of the human companion, thereby, to maintaining a good relationship with him/her.

Figure 3 illustrates the proposed HRI framework. The sensory data from various sensory interfaces are integrated in the multimodal interaction module. The integrated data are given to the cognitive interaction module and emotional

Fifteen characters per inch standard or an equivalent proportional spaced type such as "Times New Roman" font is preferred. This sample of the main text uses the 10-point "Times New Roman" font. interaction module to produce the desired action/response. The produced agenda from both modules are merged in the integrator, and the merged agenda are transferred to multimodal interaction module if the agenda are responses to the human companion, or transferred to the action part if the agenda are actions to be performed by the robot. It should be noted that even though the structure of HRI has been divided into three modules, these are closely related to each other and dynamically influence the processing in other modules, being mutually reciprocated.

III. Modules in the HRI Framework

3-1. Multi-modal Interaction Module

Multimodality does not simply mean the number of sensory input/output channels. It refers the quality of a system that allows it to perceive, represent, and interact the world using a variety of modalities in a way to enrich perceptual interaction between human and robot. In fact, multimodal interaction is so natural to human communication as found in daily life activities – i.e., hand gesture during conversation. One of the practical reasons why multimodal interaction is required is that vocal interaction alone is not always adequate to comprehend one's intention. Therefore, the other modalities, usually visual information are used as supplementary to the vocal information. As a general rule of thumb, information from a single modality alone is not adequate to guess the human intention, therefore, additional information from different modalities are used to resolve the ambiguity. On the other hand, humans use gestures in addition to verbal expressions to have their intentions and emotional states properly conveyed. The body signs such as gesture and facial expression contain rich information to indicate a person's emotional state and intention. For these reasons, a robot needs to have ability to process various sensory inputs and to use gesture and facial expression.

In the monumental experiment carried out by Bolt [8], the command "Put that there", which processes both speech and pointing gesture [8] were used. In his experiment, the human companion used voice commands together with pointing gestures to refer to a certain object and place. This experiment showed that voice can be augmented with simultaneous pointing gestures. In addition, Bischoff et al. used voice, keyboard, or e-mail for their robot 'HERMES' [9]. The transmitted inputs are combined and translated using natural language processing technology. Another robot 'Coyote' that can manipulate multimodal information has been developed by Perzanowski et al. [10]. Users can combine speech, gesture, and PDA inputs in various ways to interact with this robot.

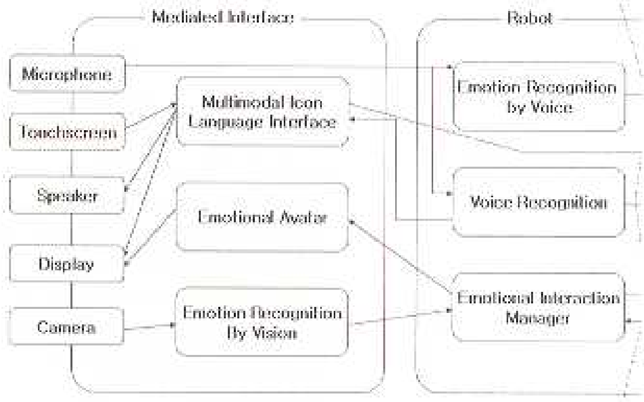

Figure 4. Configuration of Mediated Interface

There are several important issues in the multimodal interaction that are worth to be noted here. First, the inputs from different modalities are intrinsically different in terms of time scale and their specific mode. How to bind multimodal inputs in a consistent manner is a critical issue. For example, the 'red apple' indicated by voice and a finger might involve in at least 3 perceptual processes such as object, gesture, and speech recognition. The outputs from different modalities can be asynchronously generated in each own output mode. So, at a certain integration stage, each modal output is necessarily to be integrated within a limited time range and a representational form. Second, inputs from many modalities do not necessarily require for a current task. Moreover, some of them would be conflict with others. The system has to select relevant as well as consistent information. For this, our multimodal interaction module has a selective attention module in order to reduce the amount of information from multimodal inputs, and discriminate and weight an important input at an instant. In addition, one way to impose selection criteria on the incoming inputs is to use top-down knowledge of the situation. Therefore, only consistent or useful information of current situation is selected and processed. After selection, inputs can be weighted according its degree of importance, and integrated into a unified input set.

The current technologies for face, speech and gesture recognition lack stability in dealing with patterns due to their wide range of variation. For example, the speech recognition rate is sensitive to the distance from the speaker. So, we might need redundant means to use in case when the primary recognition system fails. For this purpose, we use a mediated interface. Figure 4. shows the configuration of our mediated interface. The mediated interface is portable, and it has a vocal interaction channel as well as a GUI. It also has a built-in camera, which is used to recognize facial expressions especially when the user is far from the robot. The mediated interface has more merits in addition to above rationale. Since the mediated interface would be used on a user's palm-top, a number of problems of the sensory system can be avoided. The problems due to variable distances and having to deal with too many objects are eased in the process of face recognition. It also helps to reduce the number of voices gotten into the speaker dependent voice recognition system. It would therefore, improve the efficiency of the robot as the robot does not need to get close to the user to interact.

3-2. Cognitive Module

This section presents the description of the cognitive interaction framework for HRI and its components. HRI is essentially the process of communication between human and robot. Therefore, it needs to have knowledge and dialogue skills comparable to human-human interaction. Especially HRI needs models on how the human and the robot interact, what the human does and how the human behaves, what the robot does and how the robot behaves, what the human should know, and what the robot should know.

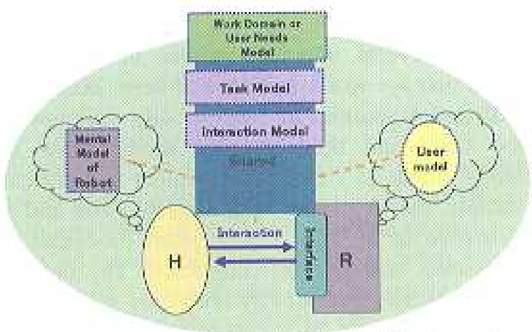

Figure 5. A conceptual framework for cognitive interaction

The conceptual illustration of the proposed framework is shown in Figure 5. It can be divided into 5 component models – needs model, task model, interaction model, user model, and mental model. Mental model is the intangible model that the humans use to describe themselves and the robot with which they interact [11]. The remaining four models form the cognitive model of the robot. The logical relations between the component models in the cognitive model of robot are shown in Figure 6.

1) Needs Model

Needs model is not related to the goal or the task, but it provides constraints which should not be violated during performing a task. For example, the robot should not make the user bother, and it should perform a sensible behavior when serving a user.

2) Task Model

Task model concerns the declarative and procedural knowledge of the task that the robot should perform. It also includes the knowledge of how the task could be performed. A task plan can be made from user's mission, object frames of the task, and the partial task scripts which are related to the properties of the task object (o). For example, a "fetch-and-carry" task can be expressed as a series of sub-tasks: Determine(o) → Move-to(o) → Get(o) → Move-to-user-with(o) → Deliver(o). That is, a partial plan can be added or deleted depending on the properties of a requested object, even though the general scheme remains the same regardless of objects. In comparison with Bring-to-user(cup), for instance, Bring-to-user(cola) may requires additional tasks such as Open-door(refrigerator) and Close-door(refrigerator) in the Get(cola) subtask.

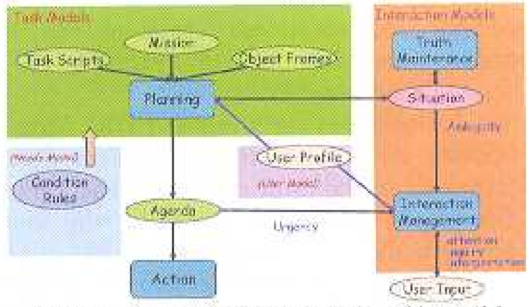

Figure 6. A schematic architecture of cognitive model

3) Interaction Model

Interaction model is responsible for the human-robot communication, and it carries out the dialogue between the human and the robot. The proposed model uses two dialogue control parameters; ambiguity and urgency. Ambiguity governs the necessity of interaction to resolve uncertainty and incompleteness in beliefs which are related to the situation and the properties of a requested object. If the robot is confronted with an ambiguous situation and it cannot perform the task, it asks the user about the situation and gets an advice. Urgency governs the time stamp of the interaction (when the interaction is initiated). It is calculated from the agenda and the current task situation. The robot can avoid bothering the user by treating only urgent queries.

4) User Model

The user model is a representation of the knowledge and preferences of the user [12]. The robot keeps this model and uses it to provide a personalized interaction to a user. The knowledge represented in the user model may be acquired implicitly during human-robot interaction, or it may be explicitly constructed by such means as interviews, questionnaires, etc.

3-3. Emotional Interaction Module

The capabilities of comprehension and expression of emotional states are fundamental to HRI, the same way as it is important in human-human relation that are described in psychological studies. Consideration of emotions is extremely important in cognitive process such as decision making, planning and learning. Emotions are also important in social interactions such as verbal and non-verbal communications. In fact, humans express emotions even with lifeless objects such as a car or a toy.

Many current researches about emotional robots are based on investigating human psychology. For instance, Kismet developed by Breazeal and her colleagues has the ability to perceive visual and auditory cues including gaze direction, facial expression, and vocal babbles, and expresses various emotional states such as happiness and sadness [4]. In a series of experiments using kismet, it has been shown that emotional interaction makes humans to interact with robots easily. However, is it really necessary that a service robot, which one might believe that it should behave like 'a good boy', has to be emotional? Hints to this question may be found in ISR (Intelligent Service Robot) of KTH that carried out 'fetch-and-carry' task in an office environment. The results of the evaluation tests using humans indicated that the emotional interaction is desirable. Similarly, the female receptionist robot "Valerie" of CMU has colorful personality and biography, and performs an attractive social interaction with the visitors [13].

The emotional interaction module places emphasis on adaptive aspects of the emotional interaction. This implies that even though a main character trait of our service robot is caring and polite, it might show different emotional expressions depending on the situation and the user. For example, our robot may show a cautious and aggressive response to an unknown person who enters into the house without being welcomed by the host.

Proposed emotional interaction model is composed of two layers - reactive and deliberative layer. Each layer is in charge of different types of emotional interaction. The reactive layer is responsible for immediate emotional reaction. Through this mechanism, a service robot can express its valence to events and release an emotionally reactive response to users. This is a very simple process governed by external (event) and internal (motivation or drive) releasers. An emotional behavior of a robot takes sensory inputs and produces corresponding emotive actions as an output, so that the emotive action is tightly linked with a perceptual state. When a robot faces a situation that makes it aroused, the robot assesses behavioral candidates based on the situation as well as internal drives. For example, a service robot would be self-motivated to contact and interact with users as many as possible. When it got bored, it becomes aroused and releases 'Contact_Human' behaviors. Also, it would maintain an equilibrium state such as keeping sufficient electric power. What we aim to achieve using this reactive mechanism is that a service robot is more likely to be 'believable' as exhibiting self-motivated behaviors. We expect that the 'close' relationship between human and robot would be constructed if a robot becomes more believable. Consequently, a service robot can be viewed as a companion rather than a lump of machine.

On the other hand, the deliberative layer supports secondary or indirect emotion that concerns with tasks and the social relationship with a user etc. Also, it tightly is coupled with the cognitive module. That is, it provides emotional states to the cognitive module and receives goals and states of a current task from the cognitive module. Our deliberative layer focuses on task-driven activities rather than social relationship with a human user. The deliberative emotion is appraised on the basis of events, agents and objects found in Ortony's model [14]. The event-related emotion comes under the category of events appraised by the appraisal variables such as desirability praiseworthiness, appealingness, etc. As evaluating status of task execution, a user's emotion and environmental condition, the robot expresses corresponding emotional states such as happiness, satisfaction, disappointment relief etc. In addition, a robot may propose an alternative task if a requested task cannot be acceptable and then produce uncomfortable emotional states.

In the proposed model, the emotional module consists of two different layers in which each layer take a different role in generating emotional expression and action simultaneously and independently. One of theoretical challenge in emotion modeling is to construct a coherent structure that accommodates both layers. In our approach, an additional mechanism so-called 'priority judgment' is required to deal any conflicts caused by two layers. So, the reactive emotional responses would be suppressed if it is not proper to carry out a task. On the other hand, the reactive layer may alarm the deliberative layer and interrupts a current task in order to deal with emergency. After this priority judgment, the selection of emotional states and expression is transmitted to further processes for controlling action or body movement.

In this way, our emotional module aims to establish a close relationship with human users. However, it should be necessary that the models will be evolved by human evaluation in daily life and real world. The iterative modification will be essentially accompanied with experiments with human subjects.

IV. RECIPROCAL INTERACTION

This section describes the reciprocal interaction of modules in order to achieve the proposed HRI framework. The proposed modules are interconnected in ways to provide the mutual influence between them. The benefits from this reciprocal interaction are expected to obtain the system stability and dynamic behavior modulation. The system stability would be achieved by resolving possible conflicts produced by modules, and the robot behavior be dynamically modulated by mutually providing their outputs to the other.

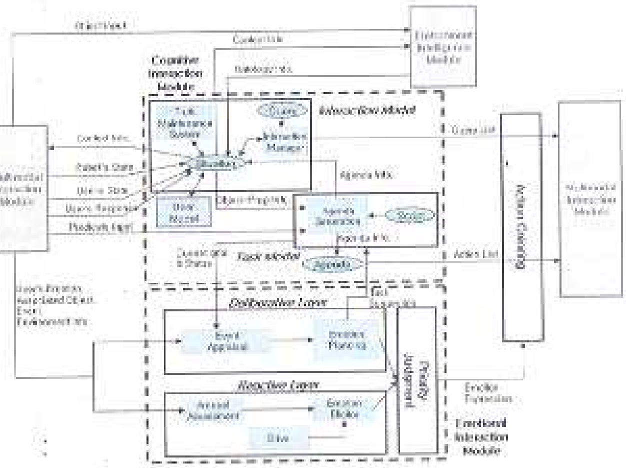

Figure 7. Reciprocal interaction between cognitive interaction module and emotional interaction module.

4-1. Reciprocity Between Cognitive and Multi-modal Interaction Module

The interaction from multimodal to cognitive interaction module concerns the resolution of ambiguity. The uncertainties are caused by sensors, manipulators, and human instructions. Integrated information from multimodal processes may not be adequate to comprehend user's intention. The ambiguity need to be resolved in order to carry out an instructed task. The ambiguity can be resolved by questioning the user, or inferring a likely candidate in a situation. On the other hand, the interaction from cognitive to multimodal interaction module can be achieved by imposing top-down knowledge on the multimodal interaction module. That is, top-down knowledge about the situation or human user can be used to determine what modal information should be selected at a certain interaction stage. In principle, the interaction between multimodal interaction module and cognitive or emotional interaction modules is similar to that between bottom-up and top-down information.

4-2. Reciprocity Between Emotional and Multi-modal Interaction Module

The multimodal interaction module is related to the emotional interaction model in two different ways. One is to recognize the human emotional state and the other is to express the robot's emotional state using face, voice tone, and gesture. The information of human emotional states can be extracted from various perceptual modalities. The extracted emotional input is classified into one of the defined emotional states such as anger, happiness, sadness etc. On the other hand, the interaction from emotional interaction module to multimodal interaction module can be considered as emotional expression at a behavior level. A corresponding emotional behavior is expressed with various action modalities as changing facial components, voice tone or making a gesture.

4-3. Reciprocity Between Emotional and Cognitive Interaction Modules

This is the most challenging task and the core of the proposed framework of HRI to make it more plausible and natural. It is widely known that emotional factors influence decision making at a cognitive processing stage. For example, an infant's cries easily draw his/her parent attention and they may believe that he/she is hungry or sleepy. Again, cognitive factors influence emotional action in a similar way.

In our HRI framework, emotional influences to cognitive process are considered in terms of 1) decision making – to decide whether a robot requests an interaction with a human user or not: a proper time point to request an interaction with human etc. 2) behavior control – to inhibit (reinforce) the behavior that produces negative (positive) emotions of a user; 3) coloring actions – to color monotonic actions with various emotions; 4) self-motivated actions – to generate an action that attracts human attention or initiate an interaction with a human. Besides, emotional factors can be used for intention recognition and generation of adaptive behaviors to users.

Similarly, cognitive factors can influence emotional factors. A proper emotional response can be decided by cognitive factors such as situation awareness. That is, a situation context provides criteria of whether possible emotional expressions or actions are proper or not. For example, greeting responses can be changed depending on the situation – host or guest, man or women, young person or elderly person, day-time, evening, or night, etc. So, the robot would show more polite and gentle greeting responses to an elderly person than those to a young person.

V. Conclusion

A new framework for human-robot interaction has been presented, which encompasses the features of natural and human-centered HRI. The multi-modal interaction has been incorporated into the new framework to make the interaction comfortable to the human user. The cognitive interaction is used to make the interaction more cooperative. The new framework is helpful to build and maintain a human-robot relationship.

We expect to use the proposed HRI framework in order to develop a more general framework for human-centered robotics.

ACKNOWLEDGEMENT

This work was supported by the Intelligent Robotics Development Program one of the 21st Century Frontier R&D Programs funded by the Ministry of Science and Technology of Korea.

Reference

[1] MIT AILab, Juan D. Velasquez: "Modeling Emotion-Based Decision-Making", Proceedings of the 1998 AAAI Fall Symposium Emotional and Intelligent: The Tangled Knot of Cognition (Technical Report FS-98-03); Orlando, FL: AAAI Press.

[2] MIT AILab, Rodney Brooks, Cynthia Breazeal, Matthew Marjanovic, Brian Scassellati and Matthew Williamson "The Cog Project: Building a Humanoid Robot", in C. Nehaniv, ed., Computation for Metaphors, Analogy and Agents, Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, 1998.

[3] MIT AILab, Breazeal, "Towards Sociable Robots", Robotics and Autonomous Systems, 42(3-4), pp. 167-175

[4] C. Breazeal, "Emotion and sociable humanoid robots", Int. J. Human-Computer Studies, vol. 59, no. 1-2, pp. 119-155, July 2003

[5] Z. Khan., "Attitude towards Intelligent Service Roots", Technical Report TRITA-NA-P9821, NADA, 1998

[6] Helge Hüttenrauch, Anders Green, Mikael Norman, Lars Oestreicher, Kerstin Severinson Eklundh, "Involving Users in the Design of a Mobile Office Robot," IEEE Transactions on Systems, Man and Cybernetics, 2003

[7] Robin R. Murphy & Erika Rogers, "Human-Robot Interaction," Final Report for DARPA/NSF Study on Human-Robot Interaction, 2001

[8] R.A. Bolt, "Put that there: and gesture at the graphics interface," Computer Graphics, Vol.14, No.3, pp262-270, 1980.

[9] R. Bischoff and V. Graefe, "Dependable Multimodal Communication and Interaction with Robotic Assistants," IEEE International Workshop on Robot and Human Interactive Communication ROMAN200, PP.300-305, 2002.

[10] D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska, "Building a Multimodal Human-Robot Interface," IEEE intelligent systems, Vol.16, No.1, PP.16-21, 2001.

[11] Donald A. Norman and Stephen W. Draper (1986). User Centered System Design. Hillsdale, New Jersey: Lawrence Erlbaum Associates

[12] Benyon, D.R. and Murray, D.M. (1993a) Applying user modeling to human-computer interaction design, Artificial Intelligence Review (6) pp 43 - 69

[13] Robo-receptionist 'Valerie', http:// www.roboceptionist.com/

[14] A. Ortony, G. L. Clore and A. Collins, "The Cognitive Structure of Emotions", Cambridge Univ. Press. 1988