AN EFFECTIVE FRAMEWORK DESIGN OF HUMAN-ROBOT INTERACTION IN THE COEXISTENT ENVIRONMENT


*Mechanical Engineering, Korea Advanced Institute of Science and Technology
**Industrial 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 an effective framework design of human-robot interaction. The framework 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 human, the cognitive interaction is effective sharing of tasks, and the emotional interaction is to maintain an intimate relationship with human. 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

1. 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 niche 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 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 implement and integrate these three modules into a single system at a software level and how one of them interacts with another. Future research directions and plans are also presented in brief.

2. HRI FRAMEWORK

2.1. Approach to 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
interprets of how HRI occurs in the context of a task domain. Human-centric design issues are mostly ignored in the conventional approach. Similarly, 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

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
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-orientated, whereas the interaction attributes in emotional module are more human orientated (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 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 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.

3. MODULES IN THE HRI FRAMEWORK

3.1. Multi-modal Interaction Module

Interaction between humans usually is through the multi-modality. 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. Usually, vocal information is accompanied with these body signs. These body signs not only provide additional information to the other party, but also attract other’s attention. 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.

There are two important issues be concerned in the multimodal interaction. The first issue is which modalities are selected to use, and how they are used. The other issue is how to integrate those multimodal inputs in a consistent manner. For the first issue, our robot uses visual, auditory, and haptic modalities. Additionally, it also uses digital inputs such as GUI buttons, and other sensors such as temperature sensor, user’s health monitoring sensor, etc., to recognize a user’s intention and state accurately. This robot can also express its intention or state using verbal expression, gesture, facial expression and visual display such as GUI. For the second issue, 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. Since the recognized result from one modality sometimes conflict with other modality(s), the selection of consistent inputs is crucial. 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 Interaction 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).

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.

3.2.1. 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.

Figure 6. A schematic architecture of cognitive model

3.2.2. 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.

3.2.3. 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 can not 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.

3.2.4. 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, body posture, 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, and a human-like avatar has been used [5]. 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.

Emotional interaction model is divided into social interaction model and user’s emotional model. First of all, the multimodal interaction module classifies emotional inputs from different modalities into basic emotional classes (anger, disgust, fear, joy, interest, sorrow and surprise) and their confidence levels. These emotional states are given to both the social interaction as well as the user’s emotional model. Social interaction model represents the relation between emotional states at a perceptual level – emotional outputs at a behavior level. Information about a specific user’s characteristics and variations in emotional interactions is stored in user’s emotion model. Therefore, learning capability is essentially required for user’s emotion model. The remaining component of emotional interaction model is the robot personality model. This model works to make a robot to look like having individuality and drive. Therefore, it influences the social interaction model and colors its emotional expressions.

Social interaction model contains the behavior patterns of emotional interactions that are socially acceptable. For example, it would contain the rule for ‘greeting’ when a robot meets a new user and greets him/her at the first time. This might make it to greet with smile and to ease the user’s tension caused by dialogue with a weird talking machine.

At present, our robot is capable of being kind and polite. However, it lacks the capability to be intimate with a user. The user emotion model can supports these additional capabilities and emotional interactions. This model contains such information as personality and the history of the personal emotions of each user. This information is accumulated while interacting with users, and the robot learns to adapt its emotional responses to specific users gradually. As a result, users will tend to consider the robot as a friend who could understand them, which leads to building a long-term relationship with the robot.

The robot cannot show rich and colorful emotional expressions with those two models because it shows similar response to similar situation and the same user according to its action rules. In order to express various emotional interaction patterns and attract users’ attention, a robot is required to have various personalities. These personalities should be programmable, and invoked by the robot to express different emotional responses depending on the situation and the user. The robot personality model also has internal motivation generator, which could work without user interaction, for a specific task. For example, it will make the robot to move from one place to another to find the user, sing a song or murmur trivial monologue when the robot feels bored. The user would think the robot as a trustworthy companion if he watches such an action. This will lead the user to have more active and patient attitude toward the robot.

All such models are composed of controllable parameters, and the models have storage for patterns of these parameters. This makes the models tractable and computational. There is a possibility that these models be too simple due to the parameterization of characteristics.
and emotional interactions. Therefore, it will be necessary that the models be developed by closely watching the human emotional interactions and then modified iteratively in the experiments with people.

4. SOFTWARE ARCHITECTURE AND RECIPROCAL INTERACTION

This section describes the interaction of modules in order to achieve the proposed HRI framework. The framework assumes a restricted home domain in that the possible human-robot interactions during a number of daily-life activities are considered. This domain limitation intends to simplify the problems arising in human-robot interaction, as well as to easily find a required interaction, which is quite possible in the restricted home domain.

One of the issues arising in the proposed framework is the representation of the scenario model. The use of symbolic expression causes problems such as combinatorial explosion of predefined tasks and poor adaptability to dynamic situations. In the proposed framework, the knowledge structure about tasks and interactions consists of record fields and a rule-base that correspond to declarative and procedure knowledge. We are concerned about two main architectures to implement and integrate systems in the framework, namely, the integrated software architecture and reciprocal interaction architecture.

4.1. Software Architecture

It is very important that each module is built by a unique method and the information at each module flow in a consistent manner. Previous works have emphasized the different characteristics of the inter-module interactions, and we intend to integrate the modules to produce a coherent behavior without losing their own properties. For this purpose, our choice for the systematic software integration is the process and component structure. Components are reusable elements that can be combined with other components, and used to construct a specific process. For an example, a face recognition process may include several components such as pre-filtering, face position detector, face recognizer, etc. Some of these components can be reused to construct other process as well. Similarly, each module is constructed by several processes. Therefore, modules or systems in our framework have a decomposable hierarchical structure. A component is defined by CORBA protocol on Linux platform that has real time capability of hardware control. Information processed by each component is prioritized by a central planner in order to constitute a series of actions and thereby generate a behavior. One of the issues along with components is the database pertaining to a class or objects recognized by each modality. Information received through various modalities should be stored in a common/shared data structure. The data structure should have a data field that allows processes to access it. For this purpose, our databases are designed by XML that has multiple data fields for different modalities.

4.2. Interactive architecture

Reciprocal interaction between modules should be considered at the software architecture level. Here, we consider 3 different types of mutual interactions between modules.

4.2.1. Reciprocity between cognitive and multimodal 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.

4.2.2. Reciprocity between emotional and multimodal 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.

4.2.3. 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 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.

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**


