A Formal Model of Sharing Grounds for Human-Robot Interaction

Jung-Hoon Hwang, KangWoo Lee, and Dong-Soo Kwon, Member, IEEE

Abstract—In human-robot interaction, interactivity is a critical issue in terms of how they understand each other and how they make other party understand efficiently. In this paper, this issue is addressed formally with respect to the mental model and shared ground. A formal model of human-robot interaction based on the information theory is presented in order to explain the mechanism of human-robot interaction and the role of shared grounds and the mental model of human-robot interaction. It is expected that the presented model of H-R interaction will increase the understanding of human-robot interaction, and may be useful in classifying, describing and predicting this interaction process.

I. INTRODUCTION

TRADITIONALLY, robots have been limited to use in helping human workers in mainly industrial and scientific areas such as assembly factories, shipyards, nuclear plants, and even the space shuttle. However, in the near future robots are expected to be popularly used in human’s daily lives. These robots for daily life will do various tasks rather than one specific task, and will interact with ordinary people rather than robot experts. Accordingly, a great deal of effort has been exerted to develop a communication and coordination system between humans and robots in order to realize an efficient, collaborative, and affective relationship. However, the majority of these efforts have put too much emphasis on the interface technology of the robot, i.e., the development of a robot interface system that implements vision and speech. Thus, little attention has been paid to what mechanism governs human-robot interaction, how human users construct a mental model of a robot, how a user’s mental model acts in terms of interacting with the robot and establishing the shared ground for efficient communication and cooperation with the robot.

The importance of interactivity becomes clear if autistic children who have difficulty in understanding other’s mind [1] are considered. Typically, these children show poor communication and social skills; thus, they are difficult to interact with - not only for their peers but also for their caregivers. By this analogy with autism, interactivity in human-robot interaction is a critical issue in terms of how they understand each other and how to make each party understood efficiently. Though this problem is tightly linked to the ‘theory of mind’ theory [2, 3], it will be approached with respect to the mental model and shared ground in a formal manner.

As noted above, human-robot interaction (HRI) is an important issue in robotics as well as in relevant academic areas, and researchers in these areas pay more attention to the idea of ‘how a robot can communicate and cooperate with a human’. Nevertheless, there seem to be few accounts of a theoretical framework that guide a designing, evaluating and explaining of human-robot interaction. Much of the current research regarding the topic is fairly technically-oriented, and thus suffers from a ‘poverty of theory’. This has not originated simply from the current limitation of artificial intelligence (A.I) technologies or from a lack formalization of psychological studies. Rather, it is because few theoretical linkages between the two agents with heterogeneous abilities have been developed thus far.

In this paper, a formal model of human-robot interaction is presented that explains how shared ground and the mental model influence the establishment of a human-robot relation. It is also shown that the presented model can be used in the expression and prediction of the interaction process. A brief introduction of the shared ground and mental model is presented in Section 2. The model of human-robot interaction is presented in Sections 3 and 4; a simple pilot experiment to verify the proposed model is presented in Section 5, and the conclusion follows.

II. SHARED GROUND AND MENTAL MODEL

Consider a simple situation where a human user asks a robot to ‘bring a cup’. In this simple statement, a vast amount of assumptions – what the robot can do, what it recognizes, and many others are implicitly made, and a proper response from the robot is driven from actions corresponding to the assumptions. That is, communication and coordination are not simple exchange processes of transmitting an utterance or action and receiving a response. Rather, they involve collective actions that are built on shared ground such as mutual knowledge, mutual beliefs, and mutual assumptions [4, 5].

The mental model and shared ground are closely linked in terms of the interaction between two agents. Mental models generally describe a person’s representation of what the external world might be or how it might work (e.g. how a computer system works) [6, 7, 8]. A mental model is not static,
rather it is dynamically evolved through interaction with the world. It can be acquired and expanded with one’s experience or interaction with the world, and constructed into various forms – knowledge, belief, assumption, etc. These representations driven from one’s experience can be shared with others if they have similar internal representations of the world. For instance, an object ‘computer’ referred to by a speaker can be easily understood if the listener has experience in dealing with it, whereas an object ‘bibim-bab’ – boiled rice and vegetables mixed with a spicy paste – can not be understood if the listener has not tasted this Korean traditional food.

In this sense, shared ground is fundamental for any interactive activities such as communication and coordination. However, this does not imply that sharing grounds is a unilateral process wherein shared ground enables individuals to communicate with each other. Rather, it is a mutually pervasive process in which shared ground can be built up through interactions between individuals, and vice versa.

III. A GENERAL SETTING OF HUMAN-ROBOT INTERACTION

In general, an interaction occurs between two independent agents A1 and A2, when Agent A1 constructs some message (or information) m, passes it to Agent A2, and m influences the behavior of Agent A2, as shown in Fig. 1.

The agents A1 and A2 may possess some knowledge that constructs or interpret the message m. The knowledge can be about the surrounding environment – i.e. objects, events, places; it can be about the other party, i.e. user identity, profile, system; or it can be about the agent himself, i.e. what he can do and how he can do. These components could be represented differently in an agent’s system (or mental model). For instance, a refrigerator can be represented as a movable object for a human, while as a fixed object for a robot. The discrepancy of knowledge representation may be caused by a qualitative and quantitative difference between two agents’ perceptual, cognitive and emotive systems.

Therefore, the shared ground plays an important role that links two heterogeneous agents. It is expected that the knowledge of each agent evolves to share grounds through interactions between the agents so as to maximize the efficiency of each interaction. Moreover, changes in knowledge are continuously occurring during interactions. In the following section, it is shown how these changes in knowledge influences shared ground and the mental model in a formal way based on information theory.

A1 \( \xrightarrow{m} \) A2

Fig. 1. Information transfer between two agents [9].

IV. FORMALIZATION OF HUMAN-ROBOT INTERACTION

A. Basic Definitions in Information Theory

Information theory is a formal way to measure the information in a signal [10]. In information theory, the measure of the uncertainty or information about a random variable is called entropy. The measure of the amount of information of a random variable that excludes the amount of uncertainty regarding another random variable is termed conditional entropy. The measure of the amount of information that one random variable contains about another is termed mutual information. If there are two random variables X, Y, the relationships are as follows:

Figure 2 shows the relationships. The measure of the information of X and Y is \( H(X) \) and \( H(Y) \). The measure of the amount of information that X contains about Y is \( I(X;Y) \), and the measure of the amount of information of X excluding the amount of uncertainty in Y is \( H(Y|X) \).

\[
\begin{align*}
H(X) & \quad H(Y) \\
H(X;Y) & \quad H(Y|X)
\end{align*}
\]

Fig. 2. Relationships between two random variables

B. 4.2 Modeling of Human-Robot Interaction with Information Theory

The interaction between two agents, human and robot, can be expressed in terms of a three way relationship among a human, robot, and the world excluding them. That is, the interaction can be considered as a communicative or mutual activity of the world or of the agents themselves. This also implies that efficient human-robot interaction should be commonly grounded with the world or with the human and the robot. Any interactivity without common ground may cause the other agent to become confused and misunderstood.

One way to formalize this three way relationship is based on information theory, in which three variables, the knowledge of humans, the knowledge of robots, and the external world, can be formalized with entropy and mutual information. Entropy denotes uncertainty of an element from an external viewpoint. Mutual information is a shared part of two or more elements and can be considered as information without uncertainty between those elements. By using these concepts, the relationships are expressed as in Fig. 3 (a).

H denotes the knowledge possessed by a human user; R denotes the knowledge possessed by robot; and E indicates external or environmental objects or events excluding humans and robots. The knowledge is composed of elements that refer to elements of H, R and E. The knowledge can be expressed as a symbolic set of propositions.
and those that are not shared. Each area may represent ground that is associated with symbolic knowledge or with the physical world. It is also closely related the patterns of interaction, as the interaction pattern is different according to which ground the agents are. As an example of this, human-robot interaction may occur efficiently at the shared area by H and R or H, R and E.

![Fig. 4. Three components of uncertainty from the R’s viewpoint.](image)

C. Uncertainty

To improve the interactivity between a human and a robot, it is necessary to remove the uncertainty caused by discrepancies in the agents’ knowledge. Fig. 4 shows uncertainty from the robot’s viewpoint. The uncertainty from the robot’s viewpoint is the area obtained by subtracting the total entropy of H(R) from H(H, E, R).

\[
H(H, E | R) = H(H, E, R) - H(R) \tag{1}
\]

The uncertainty from the robot’s viewpoint (\(H(H, E | R)\)) can be divided into three parts - H(H | R, E), H(E | H, R), and I(H | E | R). In each case, a robot may be required to remove uncertainty in order to interact with the other agent.

First, the uncertainty H(E | H, R), that is shared by neither human nor robot, is an unknown portion of the world for a robot. A robot may obtain new knowledge (I(R; E | H)) by exploring the unknown object through its learning process (or data collection process) and reduce the uncertainty of H(E | H, R). For example, a rescue robot detects an object that could be an explosive. A human operator is not also sure what it is on basis of the obtained information from the robot. In this case, a robot explores the unknown object and collects more information. For another example, a robot could learn that a glass will be broken if it hits the hard surface, when a robot sees a glass being broken in such a case.

The uncertainty H(H | R, E) is the unknown portion about a human user from the robot’s viewpoint. As in resolving the process of uncertainty about the external world, the interaction process between robot R and user H, to reduce the uncertainty of H(H | R, E) and increase the shared information of I(H | R; E), is the robot’s learning process about the user. For example, robot learns that user A is usually joyful (or depressed). However, since the information in this category is not connected to the external world, the information still remains conceptual and abstract.

The uncertainty I(H | E | R) will be the major source of uncertainty from the robot’s viewpoint. The user’s preferences are included in this category. For example, suppose that the human user prefers coffee to tea. A robot learns that the user has a strong preference for coffee through interactions. When the user asks the robot to bring a drink, coffee - rather than tea - can be more a probable object that the user is requesting. The robot can alter its behavior according to whether it knows the user’s preference. A robot can resolve this kind of uncertainty by obtaining pertinent knowledge of the world during the interaction, and this process illustrates information that is moving from I(H; E | R) to I(H; R; E).

The three processes are expressed in Fig. 5. Arrow 1 represents the robot’s learning process about the external world, arrow 2 represents the learning process about the user, and arrow 3 denotes the learning process about the relationship of the user and the external world.

![Fig. 5. Three classes of robot’s learning process.](image)

D. Human’s Uncertainty

In contrast to a robot’s uncertainty, the uncertainty from the viewpoint of user H can be expressed as H(H, E, R)−H(H)=H(E, R | H), and H(E, R | H) is also divided into three parts, H(R | H, E), H(E | H, R), and I(R; E | H). The process of resolving the uncertainty about the external world moves information from H(E | H, R) to I(H; E | R), and the process resolving uncertainty about the robot R -e.g. whether the robot can dance? – moves information from H(R | H, E) to I(H; R | E).

From the viewpoint of the human user, the majority of uncertainty is I(R; E | H) during the interaction. That is, a human user has little idea of what objects a robot knows about, how it behaves in a certain situation belongs to this uncertainty category. The human-robot interaction to resolve these uncertainties changes the state of the ground between the human and the robot from I(R; E | H) to I(H; R; E), as a human user improves his understanding about the robot. For example, a user may ask for the robot to bring a chair, even though the robot can not actually transfer it, if the user does not have the proper knowledge about the robot. Improving the user’s understanding may contribute to a more efficient interaction between the human and the robot.

The shared information I(H; R | E) needs to be investigated. As the information within this category is not directly connected with the external world, it is still vague and abstract. To be used by a user or a robot, the information needs to be more specific, as represented by the user’s repeated emotional state or by the behavior or robot’s behavior in a certain situation. Thus, the shared information I(H; R | E) between the user and the robot needs to have a grounding in the external world meaning that the information moves to I(H; R; E).
world can be explained with this model. Every interaction contains the process of establishing or using shared ground. The establishment of the shared ground during the interaction is explained in greater detail in the next section.

E. Establishment of Shared Ground

As shown in the Fig. 6, most of the interaction between a human and a robot falls within four categories: transferring a human’s grounds regarding the external world to a robot (region 2 \(\rightarrow\) region 1), transferring a robot’s grounds of the external world to a human (region 3 \(\rightarrow\) region 1), building a repository of the shared ground between a human and a robot regarding the external world (region 4 \(\rightarrow\) region 1), and already having sufficient shared ground (region 1).

![Fig. 6. Four categories of information movements during human-robot interaction.](image)

The information in region 1 of Fig. 6 is \(I \subseteq (H \cap E \cap R)\) or \(I(H; R; E)\). This region has well-developed shared ground that does not require an additional exchange of information. Interactions with the information of this region will be smooth and successful while accomplishing a desired task.

The information in region 2 requires a sharing of a human’s grounds to a robot by means of instruction or teaching. A user teaches a robot how to manipulate a glass, tells the robot a name of flower while indicating the flower, or a robot observes a user to know his preference as to what he/she listens to while it is rainy, or what he/she drinks after exercising, for example. This process can be expressed as function \(f_{hr2}\), as shown in Eq. 2.

\[
\begin{align*}
  \text{f}_{hr2}: & (H \cap E) \cap R^C \rightarrow H \cap E \cap R \quad \text{or} \\
  \text{f}_{hr2}: & I(H; E \mid R) \rightarrow I(H; R; E)
\end{align*}
\]

The information in Region 3 requires a sharing of a robot’s grounds to a user by means of informing, presenting, or describing. A robot tells a user a name of new object that a robot learned from another user, or a user watches a robot describing. A robot tells a user the name of new object that it has learned from another user. As an example, a robot tells a user A the name of a flower while indicating the flower. The information in Region 3 can be expressed as function \(f_{rh2}\), as shown in Eq. 3.

\[
\begin{align*}
  \text{f}_{rh2}: & (R \cap E) \cap H^C \rightarrow R \cap E \cap H \quad \text{or} \\
  \text{f}_{rh2}: & I(R; E \mid H) \rightarrow I(H; R; E)
\end{align*}
\]

The information in Region 4 is highly abstract shared ground between a user and a robot. For example, a user A is usually depressed in a certain situation, or a robot behaves very politely when a user is very old man. It is very difficult for a user or a robot to use the information in this region for decision making, as there is no groundings in the external world. Therefore, the shared ground in this region needs to be grounded to the external world for a successful interaction.

This process can be expressed as function \(f_{br2c}\), as shown in Eq. (4).

\[
\begin{align*}
  \text{f}_{br2c}: & (H \cap R) \cap E^C \rightarrow H \cap R \cap E \quad \text{or} \\
  \text{f}_{br2c}: & I(H; R \mid E) \rightarrow I(H; R; E)
\end{align*}
\]

The above three cases, \(f_{hr2}\), \(f_{rh2}\), and \(f_{br2c}\), require an interaction between user H and robot R to create the shared ground between H and R that is grounded in the external world.

Though this model covers many cases of human-robot interactions, here is an example of an exceptional case. Even though each agent has ground for same object in the external world, because of differences in the names or representations, an additional interaction between two agents may be required to create shared grounds. This can be expressed as

\[
I_1 \subseteq H \cap E, \quad I_2 \subseteq R \cap E
\]

where \(\text{representation}(I_1) \neq \text{representation}(I_2)\)

context(I_1) = context(I_2)

This situation can occur frequently in the course of usual interactions. Each agent refers to the same object in the external world at the physical level, but they use different symbols at the abstract level. In a human-human interaction, a human usually learns and uses multiple symbols for the same physical object or event. Thus, a robot needs to have the capability of referencing one external object or event to multiple symbols.

F. Modeling of the Interaction Process (Modeling of Mental Model Governing the Interaction Process)

To build shared ground between two agents, a user and a robot, an information transfer must occur through proper sequences of interactions. The sequences of interactions are governed by each agent’s mental model about his or her self and partner. The mental model governing the interaction also can be modeled with the proposed modeling of interaction. Using the model of interaction, each sequence of interaction can be presented, and it can show the state of the self and partner. Therefore, each agent knows what he/she has to do, and can expect to know what the partner wants.

As an example, the sequences of interactions between Agent A and Agent B, about C, are presented.

1) Agent A refers c within C during the interaction, while B does not know c.

2) Agent B informs that he/she does not know c.

3) Agent A informs c to B as it knows that B does not know c.

4) Agent B obtains knowledge about c after the presentation by A.
The utility of interaction can be measured by determining. In addition, using the modeling of the interaction actions or responses of each step of an interaction can be planned on the robot side, as the proper understanding c or not.

Though the above example shows only the mental model of pertaining to the interaction of Agent A and B, Agent A (or B) may also have a mental model of Agent B (or A)’s mental model about the interaction, thus he/she can anticipate the partner’s state. For example, Agent A showing the acknowledgement to Agent B represents that Agent A guesses that Agent B may not know whether Agent A understands c or not.

The model of interaction sequences can be represented using a state-transition diagram. Fig. 7 shows all possible state-transitions.

The model of interaction sequences can be used in the planning of interactions on the robot side, as the proper actions or responses of each step of an interaction can be determined. In addition, using the modeling of the interaction sequences, the utility of interaction can be measured by counting the steps of the interaction sequences in order to accomplish a certain interaction task.

![Figure 7. All possible state-transitions in a human robot interaction.](image)

V. PILOT EXPERIMENT

A. Design of Pilot Experiment

A pilot experiment was designed to validate the proposed interaction model. To validate the proposed interaction model, the interaction process of an information transfer between two agents was applied. In this pilot experiment, both agents were human. This has a number of advantages. The rich interaction pattern can be investigated that might be cornerstones for the investigation of HRI, and the level of information about the content of the interaction can be controlled easily.

As each agent can have different levels of information about the content of the interaction, the model was applied to the different interaction processes of two groups. One type of interaction was attained from an expert group, while the other was attained from a novice group. The interaction content of the pilot experiment was various objects that are used in mechanical engineering. For this reason, the expert group was gathered from the graduate student of mechanical engineering at KAIST. The novice group was gathered from the population of non-engineering students who attended a psychology class at Yonsei University.

In the pilot experiment, two subjects were in a chamber. A desk was placed between them. On the desk experimental materials were displayed, so both of subjects could see them. Two subjects performed different roles per each interaction unit. One subject was requested to make a request about a certain object to the other subject. The other subject was requested to follow the request. Both subjects were restricted to using only linguistic expressions. To control this condition, their faces were screened and gestures were prohibited. Fig. 8 shows the experimental environment, and Table 1 shows the various objects that were used in the experiment.

The whole process of the experiment was recorded using a camcorder.

<table>
<thead>
<tr>
<th>TABLE I VARIOUS OBJECTS FROM MECHANICAL ENGINEERING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gear box, bolt, nut, drill blade, milling blade, bevel gear, ball bearing, belt, ball-socket bearing, servo motor, dc motor, encoder, spanner, stripper, wrench, socket wrench, socket, cable tie, tab, tab handle, switch, punch, L-bracket, pulley, pliers, and other objects.</td>
</tr>
</tbody>
</table>

![Fig. 8. Scene of pilot experiment.](image)

B. The Interaction Model in Pilot Experiment

It should be noted that interactions in the pilot experiment were limited to an information transfer about the external object between the two agents, Agent A and Agent B. Information pertaining directly to Agent A (or H) or Agent B (or R) was not intended to be used in the interaction. Thus, a simplified version of interaction model was applied in this pilot experiment as shown in Fig. 9. This simplified model has a state transition probability matrix that is shown in Eq. (6). The state transition probability matrix can be used as useful information featuring the observed interaction process.

![Fig. 9. Simplified version of interaction model.](image)
C. The procedure of the Experiment

Each stage of the experiment was as follows:
1. A general instruction about the experiment is given to subjects.
2. The subjects are asked to wear a mask to screen their face.
3. Subject A sees the brief information of the name, function and photo of an object that he will ask about to subject B.
4. Subject A informs subject B about the object using linguistic expressions only.
5. Subject B looks for the object described by the subject A. Subject B is allowed to request from subject A more information.
6. If subject B selects the right object, subject A finishes the interaction process by pressing the button.
7. Subject A and subject B repeat the processes in 3–6 above by changing the role during each trial.
8. After ten trials, subjects were interviewed.

D. The Experimental Results

Ten sets of interaction processes are acquired from the two groups. Each set of interaction processes is composed of tens of interaction steps. Every state of each interaction step is classified under an operational definition by an experimenter. Following this, the classified states are applied to the simplified interaction model in order to obtain the state transition probability matrices for both groups.

The state transition probability matrices (Eq. (7) and Eq. (8)) show the differences between the two groups. In the expert group, the state mainly stays in the ‘shared ground’ state, while the state for the novice group mainly stays in the ‘having ground by oneself’ state.

The state transition probability matrix is expected to be used in measuring or expecting the interaction, hence the results of the transition probability matrix is now under study.

\[
P_{\text{novice}} = \begin{bmatrix} 0.0 & 3.5 & 3.0 & 0.0 \\ 0.0 & 34.0 & 0.0 & 7.9 \\ 0.0 & 0.0 & 33.7 & 7.8 \\ 0.0 & 0.7 & 1.4 & 8.1 \end{bmatrix}, \quad P_{\text{expert}} = \begin{bmatrix} 0.0 & 1.7 & 4.2 & 0.0 \\ 0.0 & 9.8 & 0.0 & 7.9 \\ 0.0 & 0.0 & 26.7 & 15.7 \\ 0.0 & 2.5 & 3.1 & 28.4 \end{bmatrix}
\] (7), (8)

VI. CONCLUSION

In this paper a formal model of human-robot interaction is developed based on information theory, and attempts are made to explain the dynamics of interactions in terms of a three-way relationship. In addition, the role of shared ground and the mental model in the human-robot interaction was explored, as the interactivities were classified in terms of what grounds the three variables of H, R, E share. Throughout the formalization, patterns of interactivity were discovered that might occur in actual human-robot interactions. First, any uncertainty caused by a discrepancy between the two knowledge systems is required to be resolved in order to maximize task performance. Second, for an efficient interaction, shared ground at multiple levels rather than at a single level between the two knowledge systems is required.

A formal model of human-robot interaction could describe the sequence of interaction processes and made it possible to predict the interaction process. As this model can be modeled in state transition diagram, it may also be used as a computational model. Moreover, the importance of shared ground and the related mental model was reconfirmed by the model of information transfer from a robot to a user.

The interaction model was verified in a simple way through a pilot experiment. The pilot experiment showed the potentiality of a state transition probability matrix as a measure of interaction. The metric to measure interactivity using the proposed interaction model is being researched.

To use the interaction model in actual robotic applications, the state classification should be obtained. Attempts are being made by the authors to implement the state classification using various classification methods.

It is expected that the presented model of H-R interaction will increase the understanding of human-robot interactions, the role of shared ground and the mental model. It is also expected that the model will be useful in describing and predicting the interaction process as well as be useful in measuring the interactivity.

REFERENCES