Mixed-Initiative Human–Robot Interaction Using Hierarchical Bayesian Networks

Jin-Hyuk Hong, Youn-Suk Song, and Sung-Bae Cho, *Senior Member, IEEE*

***Abstract*—As the usage of service robots becomes more sophisticated, direct communication by means of human language is required to increase the efficiency of their performance. In natural speech interaction, however, people often omit some words and rely on background knowledge or the context, resulting in ambiguity. In order to develop smarter service robots, therefore, managing the context of interaction is essential. In this correspondence, we have investigated the mixed-initiative interaction that prompts for missing information and clarifies ambiguous statements based on hierarchically designed Bayesian networks. Simulation with the Kephera II robot and a usability test have demonstrated the usefulness of the proposed method.**

***Index Terms*—Bayesian network, mixed-initiative interaction, service robot.**

I. Introduction

Human–robot interaction (HRI) is an emerging field of autonomous robotics, which is tightly connected with human–computer interaction (HCI). Since traditional autonomous agents were directed by a human supervisor and their autonomy was regarded as more important, they have only managed basic interaction with humans [1]. With the recent emphasis on the usage at homes and offices, however, autonomous robots are required to hold daily-life interactions with the human. They reside in human environments and interact with people, providing services such as delivery and notifications [2].

As communication is crucial for robots to understand the user's intentions [3], natural language, which is one of the most prominent human activities, is recognized as a promising method of communication between humans and robots [4], [5]. Thus, dialogue management, which manages the natural language interaction, has been actively investigated in the field of HCI, and it also holds great potential in the field of HRI [6], [7].

In conversation, however, people often use ambiguous expressions like "turn the light OFF" which may mean "turn OFF the light in room B." As in this case, background knowledge or the context of dialogue is often presupposed so that missing or spurious sentences appear frequently [8], [9]. In order to resolve ambiguities and uncertainties, the mixed-initiative (MI) approach has been presented in the field of HCI [10]. Contrary to a tight interaction mechanism, the MI approach attempts to solve a problem by allowing the human and the system to collaborate in incremental stages [11].

Among many issues in the dialogue management, we have particularly focused on the ambiguous expressions and tried to establish a natural communication mechanism between humans and robots, which is based on a "perfect" speech recognizer. Here, the concept of MI interaction uses the hierarchical Bayesian network model. A small home environment has been designed to verify the proposed MI-HRI management, and the usability test has been conducted to show its usefulness.

This correspondence is organized as follows. In Section II, we survey the related work. Section III focuses on the MI-HRI management by using the hierarchical Bayesian networks. In Section IV, we describe a service robot with the proposed MI-HRI management given in Section III. In Section V, we report on experimental results of the proposed method in comparison with the conventional approaches. Finally, the conclusion is given in Section VI.

II. Service Robot Interaction

The service robot is an important application area for autonomous robots, because it resides in human environments such as homes and offices to interact and cooperate with people [12]. So far, there have been many studies on service robots which can inform humans about home services, deliver things, or guide people [2], [13]. However, traditional studies on service robots have mainly focused on autonomous behavior rather than interaction ability [2]. Accordingly, even a well-performing robot could not understand what a user really wanted due to unsophisticated and unidirectional interaction.

Although there are many ways to interact with robots such as menus and keywords, a survey conducted by Hüttenrauch *et al.* [12] showed that 82% of the participants in this correspondence preferred speech to other interaction media. Thus, various robots have been developed to interact with people by using natural language, but most of them could only understand simple commands rather than complex dialogues [2]. Some sophisticated language models in the robot area have been proposed. Lauria *et al.* [3] used natural language to give directions or to get robots to understand these directions. This model incrementally accumulated knowledge about paths through interaction with humans. After a robot learns a path, the user can ask the robot to deliver something to the location. Skubic *et al.* [14] proposed a sophisticated dialogue management system for spatial relationships of environments such as homes or offices. Since this kind of robot should understand conversation using spatial information and extract spatial relationships, they proposed a spatial language model that is programmed to be integrated into a robot. Lemon *et al.* [15] focused on dialogue management for collaborative activities of mobile robots, involving multiple concurrent tasks. They used a dialogue move tree and an activity tree, and supported multiple interleaved threads of dialogue about different activities and their execution status.

Since dialogue plays an important role in HCI and speech communication systems, various dialogue management techniques have already been investigated in the field of HCI: pattern matching techniques, canned scripted models, frame-based models, finite-state models, and plan-based models [16]–[20]. Many commercial systems were produced to provide information, to make reservations, to educate, and to guide web sites [17], [20], [21].

Two types of conventional dialogue modeling are system-initiative models and user-initiative models. The former has a complete control in providing a procedural guidance, while in the latter, the user has the control to determine the preferred course of interaction. Since both of them have limitations in task completion, a compromise between those two approaches might be effective. For expressions and situations hard for these approaches to deal with, the MI model has been suggested [8], [22]. Meng *et al.* [8] proposed an MI model using belief networks and a backward inference. Bechet *et al.* [23] attempted to build a two-step process and bridge a gap between the speech recognition and the language understanding. A hierarchical Bayesian network approach

Manuscript received August 15, 2005; revised October 1, 2006. This work was supported by MIC, Korea under ITRC IITA-2006-(C1090-0603-0046). This paper was recommended by Associate Editor A. Maciejewski.

J.-H. Hong and S.-B. Cho are with the Department of Computer Science, Yonsei University, Seoul 120-749, Korea (e-mail: hjinh@sclab.yonsei.ac.kr; sbcho@cs.yonsei.ac.kr).

Y.-S. Song is with LG Electronics, Seoul 150-721, Korea (e-mail: corlary@sclab.yonsei.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2007.906570

TABLE I
DESCRIPTION OF THE HIERARCHICAL BAYESIAN NETWORK

Category	**Level**	**Object**	**Example**
Goal (G_i)	Service goal level	Target function or service	Service function ("open-window"), service target ("open-window-room F")
Sub-goal (SG_j)	Mixed-initiative level	Mixed-initiative problem solving	Function ("open"), object ("window"), target ("room F")
Primitives (Pr_k)	Semantic primitive level	Evidence for inference	Verbal semantic primitive ("open the window"), contextual semantic primitive ("windows of room A and F are closed")

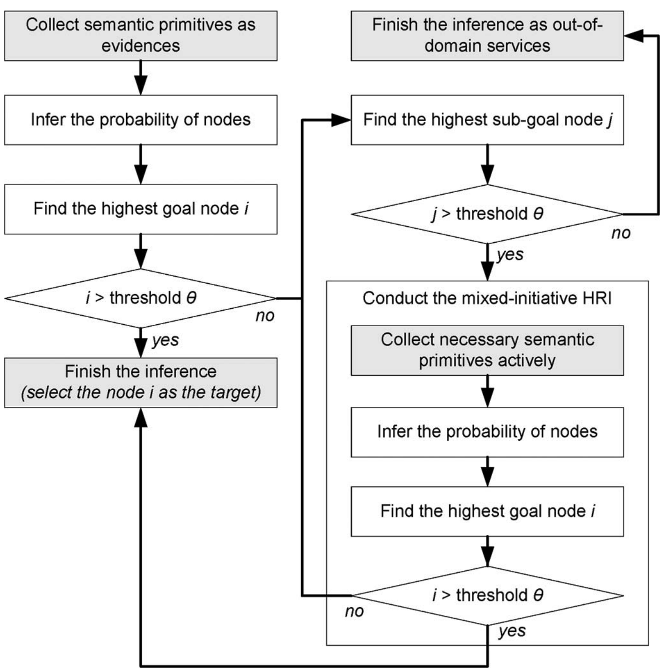


Fig. 1. Process of the MI-HRI management.

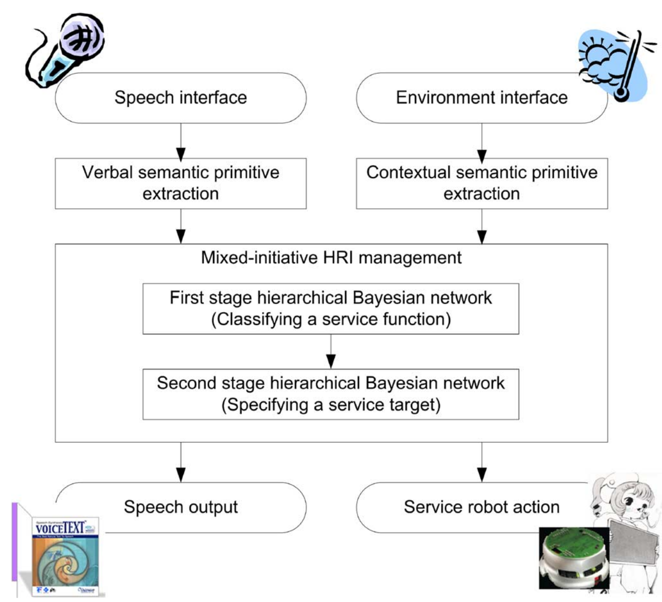


Fig. 2. Overview of the proposed service robot.

has been tried to model the context of a dialogue and to manage ambiguous natural language expressions [20]. As the domain becomes more complex, it might be difficult to infer the user intention at one try. In such a case, MI approaches might be useful to correctly infer the intention.

III. Hierarchical Bayesian Network for the MI-HRI Management

Sometimes, a query is not enough to analyze the user's intention, so the hierarchical Bayesian network is proposed for the MI-HRI management. The hierarchical Bayesian network is composed of three layers: a service goal, an MI, and a semantic primitive level. A node in the service goal level indicates a specific service requested by the user, and the MI level consists of the subconcepts corresponding to the specific service. Semantic primitives in the semantic primitive level are the pieces of information relevant to the application such as certain words and environmental variables. Since the network works within a restricted application domain, there are finite L semantic primitives, M subgoals, and N service goals (L: the number of semantic primitives, M: the number of subgoals, and N: the number of service goals). Table I summarizes the formal description of the domain and the network.

In particular, for engaging MI interaction, we positioned the subgoals between the semantic primitives and the service goals. The subgoal layer is developed for the stepwise inference. A service goal is inferred using subgoals, while subgoals are inferred using semantic primitives as evidence. When it is difficult to infer a service goal, subgoals are estimated in advance, and then, service goals are evaluated through MI interactions. Each subgoal represents a missing value, where the corresponding question predefined at the knowledge base will be provided to the user when it needs some more information for inference. We assume that there is no direct link between the goal and the primitive nodes, which seems to be similar to hierarchically constructed naive Bayes formulation.

The hierarchical Bayesian network operates, as shown in Fig. 1. All nodes of the network have probabilities between zero and one, and the semantic primitive Pr_i is true when it is included in the user's utterance or observed in the environment. With the semantic primitives, the hierarchical Bayesian network infers the *posteriori* probabilities of each subgoal and goal: $P(\mathrm{SG}|\,\mathrm{Pr})$ and $P(G|\mathrm{SG})$. Goal inference is conducted in two steps: subgoal inference and service goal inference. The former uses semantic primitives as evidence to calculate the probabilities of the subgoal nodes, while the latter puts subgoals to be used as evidence to estimate the probabilities of the goal nodes, which is shown at the bottom of the next page.

It finds out a goal node G_j with the highest probability $P(G_j|\mathrm{SG})$ that is greater than a threshold θ. If there is no goal node that satisfies the condition, however, the MI-HRI management is conducted to figure out the user's correct needs by requesting the user to offer a piece of extra information based on the subgoal level. It attempts to search a subgoal SG_k that is the greatest among subgoals and greater than the threshold θ. If there is a proper subgoal, the scope of service goals gets to be reduced. Therefore, it can actively gather necessary information to manage some missing concepts and finally complete the goal inference and determine the service. Further, basic formula on inferring intention based on the user's query might be referred from [8], [20], and [22]. Some queries, which are not manageable even with the MI-HRI management, are regarded as out-of-domain (OOD) services [8].

IV. Service Robot With the MI-HRI Management

We designed a service robot that supports the MI-HRI management, as shown in Fig. 2. The speech and environment interfaces collect semantic primitives such as "turn the light ON (verbal semantic primitive)" and "dark (contextual semantic primitive)" for the MI-HRI management. The service robot first classifies the service function using the first-stage hierarchical Bayesian network and, then, specifies the service target using the second-stage hierarchical Bayesian network. Finally, a sentence or services like "turn the light ON" might be delivered to the user after the interaction.

The speech interface supports a natural interaction between humans and robots, while the environmental interface consists of several sensors to perceive information on environment. At the stage of semantic primitive extraction, verbal primitives are extracted from the user's query by stemming and pattern matching with a predefined keyword dictionary, while contextual primitives are obtained from the sensors' observations. These values are used as the input of MI-HRI management, and speech outputs such as questions and answers are provided to the user as well as a service that the user wants as the results of interaction.

Since the purpose of the study is to investigate the advanced interaction between humans and robots, some commercial modules are employed in the proposed service robot system. "Voiceware,"[1] which is a solution for speech recognition and generation, is used to provide users with a realistic and convenient interface. A set of sensors "LabPro"[2] developed by Vernier is also incorporated to collect environmental information such as room temperature, illumination, and noises. The "Khepera II"[3] robot is used to provide services in the real application, even if the services are a little simpler than those of the simulation constructed in this correspondence.

[1]http://www.voiceware.co.kr
[2]http://www.vernier.com
[3]http://www.k-team.com

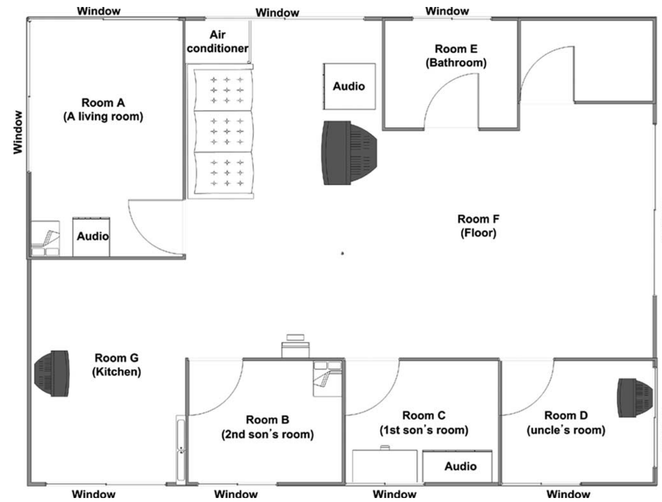


Fig. 3. Home environment implemented for the demonstration of the proposed method (home activities: Turning ON/OFF the light, opening/closing windows, and turning ON/OFF audios, TVs, and air conditioner).

TABLE II
DESCRIPTION OF THE FIRST BAYESIAN NETWORK IN CLASSIFYING A SERVICE FUNCTION

Category	Content
Primitives	Turn, on, off, open, close, light, window, audio, TV, air conditioner, cold, hot, noise, here, inside, outside, dark, current_temperature, current_luminous_intensity, current_noise_level
Sub-goals	Turn-on, turn-off, open, close, light, window, audio, TV, air conditioner
Goals	Turn-on-light, turn-off-light, open-window, close-window, turn-on-audio, turn-off-audio, turn-on-TV, turn-off-TV, turn-on-air conditioner, turn-off-air conditioner

A. Home-Service Environment

A home environment is employed to construct a simulation miniature, as shown in Fig. 3. The home services are basically initiated by the instruction of a user. The service robot extracts useful words from the query as evidence to infer services. It also observes various attributes such as temperature and illumination to model the context of environment. These values, which are used together with words (extracted to infer a service) as semantic primitives, are collected by "LabPro" which is programmed into the service robot.

$$P(G_i=1|\mathrm{SG}) = \frac{P(\mathrm{SG}|G_i=1)P(G_i=1)}{P(\mathrm{SG})} = \frac{\prod_{k=1}^{M} P(\mathrm{SG}_k|G_i=1)P(G_i=1)}{\prod_{k=1}^{M} P(\mathrm{SG}_k|G_i=0)P(G_i=0) + \prod_{k=1}^{M} P(\mathrm{SG}_k|G_i=1)P(G_i=1)}$$

$$P(\mathrm{SG}_i=1|\,\mathrm{Pr}) = \frac{P(\mathrm{Pr}\,|\mathrm{SG}_i=1)P(\mathrm{SG}_i=1)}{P(\mathrm{Pr})} = \frac{\prod_{k=1}^{L} P(P_k|\mathrm{SG}_i=1)P(\mathrm{SG}_i=1)}{\prod_{k=1}^{L} P(\mathrm{Pr}_k\,|\mathrm{SG}_i=0)P(\mathrm{SG}_i=0) + \prod_{k=1}^{L} P(\mathrm{Pr}_k\,|\mathrm{SG}_i=1)P(\mathrm{SG}_i=1)}$$

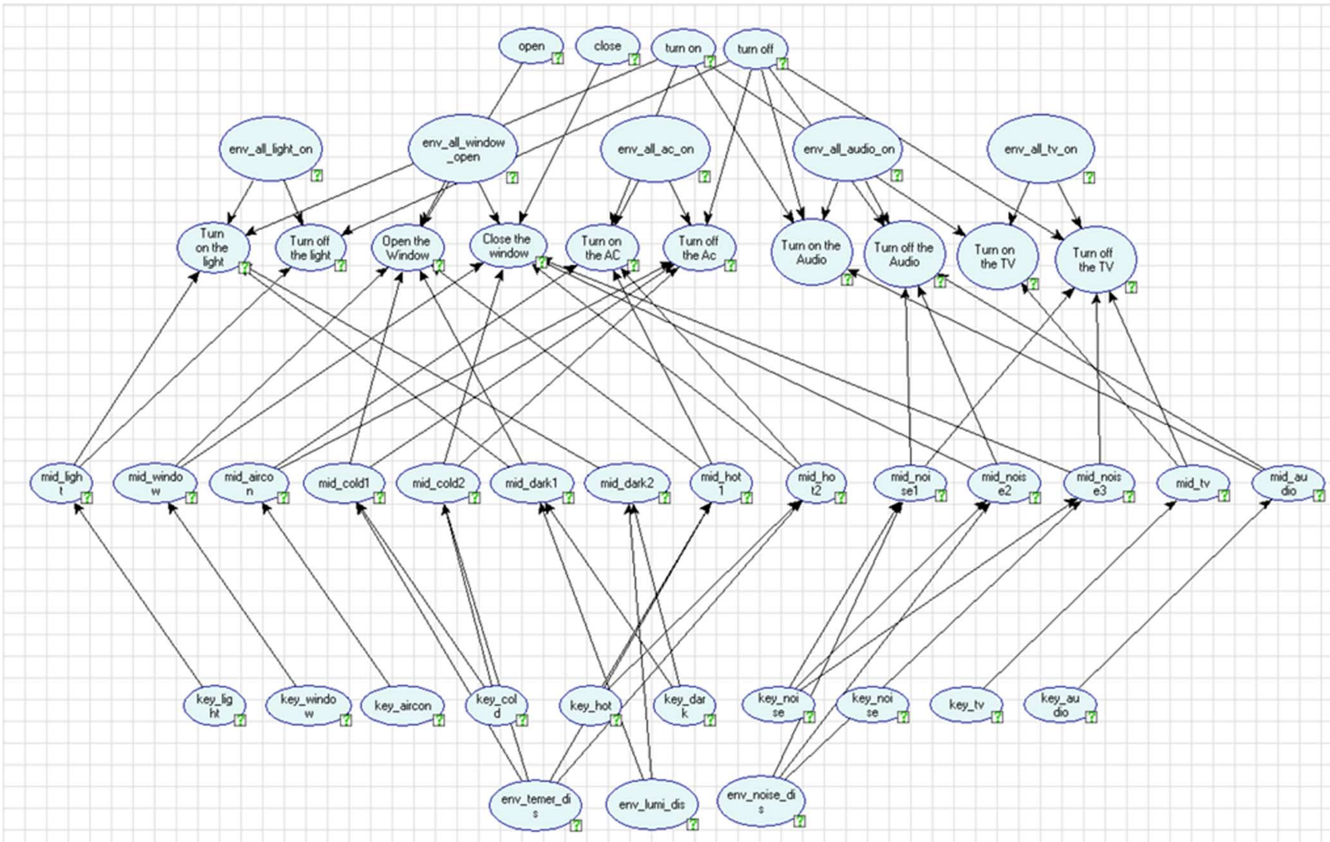


Fig. 4. Hierarchical Bayesian network for the inference of the service function.

B. Service Decision Based on the MI-HRI Management

A service robot determines a service in two stages: classifying a service function and specifying a service target. The proposed two-level service inference might increase the applicability and scalability of the proposed method, since networks in the inference model can be independently constructed from each other. In the hierarchical Bayesian network of the first stage, ten-goal service nodes are constructed by the domain expert for the associated ten functions of the service robot, as shown in Table II. A number of verbal and contextual semantic primitives consist of primitive nodes in the network, while subgoals are designed for the MI interaction. Table II describes the details of the Bayesian network in classifying a service function, and Fig. 4 shows its brief overview.

When the robot gets a command like "turn ON the air conditioner" from the user, it assigns "true" to nodes for the semantic primitives observed (turn, ON, air conditioner), while "false" is assigned to nodes for the semantic primitives not observed. Based on the inference algorithm using the conditional probability table, the probabilities of subgoals and service goals are obtained, and hence, the service goal "turn-ON-air conditioner" will be selected as the service function.

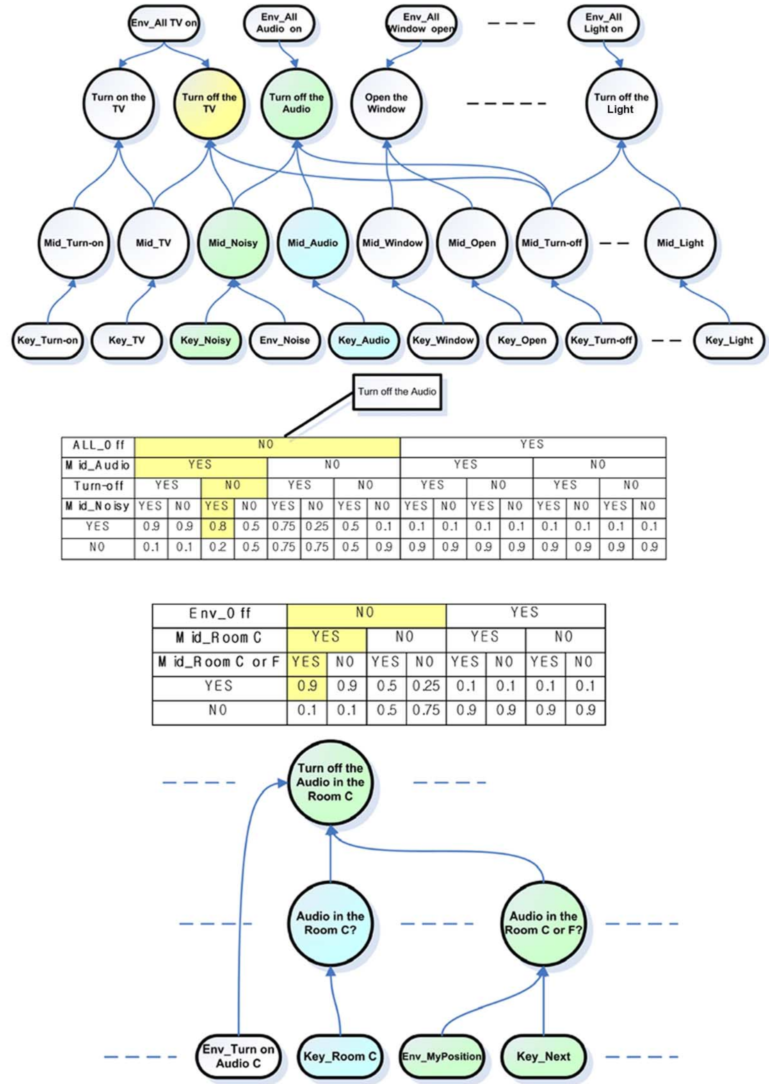
If an explicit query is given at this stage such as "please, turn ON the air conditioner," "turn OFF the light," and "open the window," the type of the service is easily inferred from the network: "turn-ON-air conditioner," "turn-OFF-light," and "open-window." If there appears an ambiguous query, however, it might be difficult to directly determine the service type. For example, "too hot here," "I am bored to death," and "turn it ON" may be interpreted as having more than one meaning. According to the specification of the home environment implemented in this correspondence, "too hot here" may signify either "please, turn ON the air conditioner" or "please, open the window." In addition, the meaning of "turn it ON" can be different according to what "it" really implies. Based on the MI-HRI management, the robot infers that the service type should be clarified and requests additional information to the user like this: "Which one do you want, open the window or turn ON the air conditioner?" The goal inference at the first stage will be completed with responses from the user. An ambiguous statement like "too hot here" might be interpreted as wanting to have an ice cream or a cold shower instead of turning ON the air conditioner or opening the window, but the set of possible inferences is basically designed based on an extendable home environment.

After selecting the service function, the associated hierarchical Bayesian network of the second stage works to specify the service target, and it decides the final service that will be performed. For example, "which room" or "which window" is determined in this stage. In this correspondence, ten hierarchical Bayesian networks, as a whole, are constructed for the second stage to determine the service target, which correspond to the service functions at the first stage. Each network also consists of several verbal semantic primitives and contextual semantic primitives related to the specific service target.

The second-stage inference is simultaneously done with the first-stage inference, and even a query that is nonproblematic at the first stage may have some ambiguities at the second stage. The command "turn OFF the light" can be interpreted as having several possibilities like "turn OFF the light in room A," "turn OFF the light in the kitchen," or "turn OFF all the lights in this house." When the user commands "power on the stereo in the next room," for other instance, the robot should know where the next room is. This might be solved by

(a)

(c)



ALL_Off	NO								YES							
Mid_Audio	YES				NO				YES				NO			
Turn-off	YES		NO		YES		NO		YES		NO		YES		NO	
Mid_Noisy	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
YES	0.9	0.9	0.8	0.5	0.75	0.25	0.5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
NO	0.1	0.1	0.2	0.5	0.75	0.75	0.5	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

Env_Off	NO				YES			
Mid_Room C	YES		NO		YES		NO	
Mid_Room C or F	YES	NO	YES	NO	YES	NO	YES	NO
YES	0.9	0.9	0.5	0.25	0.1	0.1	0.1	0.1
NO	0.1	0.1	0.5	0.75	0.9	0.9	0.9	0.9

(b)

Fig. 5. Snapshots for the action "turn-OFF-the-audio-in-room C."

using the context of the situation, but if there are several possible rooms, the robot can ask to specify the target of the service. The MI-HRI management at this stage attempts to specify the object of the service to understand the user's intention like: "Turn OFF the light in room C." Accordingly, the service will be decided and executed, or failed.

V. Experimental Results

A. Experimental Environments

A simulation using "Khepera II" robot has been made in which the environment is similar to that presented in Section IV. There are seven specific rooms, including a kitchen, a living room, and a floor; three stereos, three TVs, and an air conditioner are located in the rooms. Every room has its own windows and lights. The miniature consists of a main desktop, a model house, a microphone, a set of sensors, and a "Kephera II" robot, which moves along predefined paths. Some services, such as opening windows and turning ON the light, are considered to be completed when it reaches to the target location.

B. Illustration on Providing Services

Dialogue 1 illustrates an example using the MI-HRI management. The user wants to make the service robot turn OFF the audio in room C but does not definitely specify the task. Through the MI-HRI management based on the hierarchical Bayesian network, the robot collects more information on the user's intention and finally achieves the goal, as shown in Fig. 5. This kind of ambiguous queries that have more than one interpretation can also be handled by the proposed method.

Dialogue 1. An example for the ambiguous service function inference.

User: It is too noisy here.
Robot: Which do you want, turn OFF the TV or the stereo?
Semantic primitives (noisy, here)
Service function inference (subgoal: turn OFF) → MI dialogue
User: The stereo in the next room.
Robot: Which room? First son's room (room C) or living room (room F)?
Semantic primitives (stereo, next, room)
Service target inference (subgoal: some) → MI dialogue

Context information:

- *The user is at the sofa in the living room (room F)*
- *TV in the living room (room F) is off*
- *Air conditioner on the floor (room F) is working on*
- *Some windows are open*

Task:

1. *You want the robot to turn on the TV*
2. *You can't hear the television because of some music*

   Additional context: The music is heard from a room
3. *You feel quite hot inside*

   Additional context: Outside is hotter than inside

Fig. 6. Description on the task.

Context (stereo, room C, ON)
Context (stereo, room F, ON)
User: First son's room (room C), please.
Robot: I am going to turn OFF the stereo in the first son's room (room C).
Semantic primitives (stereo, room C)
Service function inference (turn-OFF-stereo); service target inference (room C)
(Service: Turn-OFF-stereo-room C)

C. Usability Test

A usability test is conducted to evaluate the quantitative performance of the proposed method, in which several tasks are given to subjects. The subjects are classified into the expert and novice group based on their personal experiences on the computer. A set of tasks is sorted by the degree of ambiguity of the situation. The number of interactions and the hitting rate are recorded with the log of user dialogues.

1) Task and Environment: Three tasks shown in Fig. 6 are given to 16 people (eight experts and eight novices) with two robots, where one robot uses a simple dialogue management without the MI interaction (see dialogue 2) [24] and the other adopts the proposed MI HRI. It includes the structure of the house, the current position of the user, and the condition of each object, where the service robot is set always besides the user. The manual on operating the service robot that includes several example dialogues is only provided to the expert group, while a simple hint, like that we can interact with the robot using conversation, is provided to the novice group. Before the experiment, a chance to be familiar with the robot is given only to the expert group.

Dialogue 2. An example of simple dialogue management.
User: It is too noisy here.
Robot: I cannot understand. Please say in detail.
Semantic primitives (noisy, here)
(Service: Nothing over threshold)
User: Turn OFF the audio.
Robot: I cannot understand. Please say in detail.
Semantic primitives (turn, OFF, stereo)
(Service: Nothing over threshold)
User: Turn OFF the audio in the first son's room (room C).
Robot: I am going to turn OFF the stereo in the first son's room (room C).
Semantic primitives (turn, OFF, stereo, first son room)
(Service: Turn-OFF-stereo-room C)

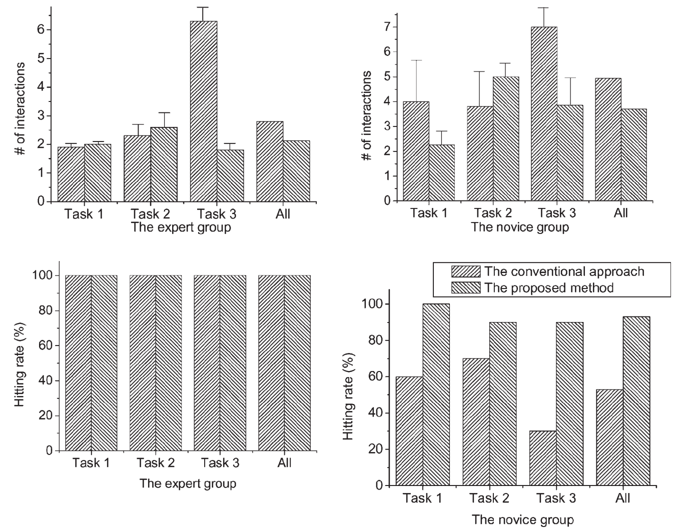


Fig. 7. Results of the usability test.

TABLE III
PAIRED t-TEST RESULT

Task	Experts	Novices
1	0.000482268	0.046570383
2	0.37535689	0.488809052
3	0.001496543	0.006331119

2) Result of the Usability Test: Fig. 7 shows the number of interactions and the hitting rate for each approach. Both the expert and the novice groups use fewer interactions when they use the proposed method. In particular, the proposed method requires less interaction than the other for the highly ambiguous situation like the third task. The subjects simply expressed their emotion like "It is quite warm here," and the service robot understood the intention of the user and suggested a solution.

In terms of hitting rate, the proposed MI HRI is better than the other for the novice group. The expert group already knew how to operate the robot, so all tasks were completely solved. For the novice group, on the other hand, it was hard to complete the given task with the conventional method. When using the conventional method, 30% of success rate was obtained for the third task. However, when using the proposed method, a success rate of 90% was achieved. The result indicates that the proposed method improves the performance of interactions between the novices and the robots, since it guides them to fulfill their goals.

A paired t-test is also conducted for each task, where we can confirm the significance level of the experimental results, as shown in Table III, except for the second task. Since we did not give any information on the source of the music, the second task might be intuitively solved to be dependent on the subject and produced rather large variation.

We have also compared the two approaches in terms of "out of domain" and "in domain" that are popular measures in evaluating the dialogue system. A query is regarded as "out of domain" if its content is not included in the domain like "who is in the room," and "in domain" otherwise. As shown in Fig. 8, there appear fewer "out of domain" queries with the proposed method than the conventional method, and the ratio of "out of domain" queries is reduced from 46.3% to 23% within a statistical significance level by using the MI HRI. It is because the proposed method restricts the scope of the conversation.

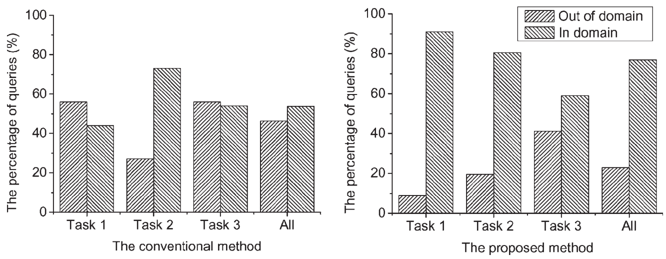


Fig. 8. Analysis of the query.

VI. Conclusion

This correspondence aimed to construct an effective conversational interface for HRI that would manage more flexible and natural dialogues. Understanding the user's intentions is important for robots to offer an appropriate service, but in conversation, it is often hard to infer the correct and detailed intention because of the uncertainty inherent in queries. Since missing concepts cause an ambiguous situation, we have proposed the hierarchical Bayesian networks to deal with it. We have also constructed a service robot working with the MI-HRI management and verified its usefulness through some simulations and usability tests.

Even though we have dealt with several issues of the MI interaction, the proposed method is limited in its applicability and robustness associated with the language understanding since the Bayesian network should be constructed for the target application. In the next stage of our research, we will investigate an automatic method in constructing the Bayesian network for a specific domain and a sophisticated dialogue model that manages various expressions of conversation.

References

[1] C. Breazeal, "Social interactions in HRI: The robot view," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 181–186, May 2004.

[2] A. Agah, "Human interactions with intelligent systems: Research taxonomy," *Comput. Elect. Eng.*, vol. 27, no. 1, pp. 71–107, Nov. 2001.

[3] S. Lauria, G. Bugmann, T. Kyriacou, J. Bos, and E. Klein, "Training personal robots using natural language instruction," *IEEE Intell. Syst.*, vol. 16, no. 5, pp. 38–45, Sep./Oct. 2001.

[4] A. Green and K. Eklundh, "Designing for learnability in human–robot communication," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 644–650, Aug. 2003.

[5] T. Fong, C. Thorpe, and C. Baur, "Robot, asker of questions," *Robot. Auton. Syst.*, vol. 42, no. 3/4, pp. 235–243, Mar. 2003.

[6] H. Prendinger and M. Ishizuka, "Let's talk! Socially intelligent agents for language conversation training," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 31, no. 5, pp. 465–471, Sep. 2001.

[7] D. Sanford and J. Roach, "A theory of dialogue structures to help manage human–computer interaction," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 4, pp. 567–574, Jul./Aug. 1988.

[8] H. Meng, C. Wai, and R. Pieraccini, "The use of belief networks for mixed-initiative dialog modeling," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 757–773, Nov. 2003.

[9] C. Sammut, "Managing context in a conversational agent," *Electron. Trans. Artif. Intell.*, vol. 5, pp. 189–202, 2001.

[10] J. Allen, "Mixed initiative interaction," *IEEE Intell. Syst.*, vol. 15, no. 4, pp. 14–23, Sep./Oct. 1999.

[11] E. Horvitz, "Uncertainty, action, and interaction: In pursuit of mixed-initiative computing," *IEEE Intell. Syst.*, vol. 14, no. 5, pp. 17–20, 1999.

[12] H. Hüttenrauch, A. Green, M. Norman, L. Oestreicher, and K. Eklundh, "Involving users in the design of a mobile office robot," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 113–124, May 2004.

[13] F. Mizoguchi, H. Nishiyama, H. Ohwada, and H. Hiraishi, "Smart office robot collaboration based on multi-agent programming," *Artif. Intell.*, vol. 114, no. 1/2, pp. 57–94, Oct. 1999.

[14] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock, "Spatial language for human–robot dialogs," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 154–167, May 2004.

[15] O. Lemon, A. Gruenstein, and S. Peters, "Collaborative activities and multi-tasking in dialogue systems—Towards natural dialogue with robots," *Traitement Automatique des Langues (TAL)*, vol. 43, no. 2, pp. 131–154, 2002.

[16] E. Horvitz and T. Paek, "A computational architecture for conversation," in *Proc. 7th Int. Conf. User Modeling*, 1999, pp. 201–210.

[17] V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T. Hazen, and L. Hetherington, "JUPITER: A telephone-based conversational interface for weather information," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 1, pp. 85–96, Jan. 2000.

[18] F. Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. Carolis, "From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent," *Int. J. Human-Comput. Studies*, vol. 59, no. 1/2, pp. 81–118, Jul. 2003.

[19] J. Williams and S. Young, "Partially observable Markov decision processes for spoken dialog systems," *Comput. Speech Language*, vol. 21, no. 2, pp. 231–422, Apr. 2007.

[20] J.-H. Hong and S.-B. Cho, "A two-stage Bayesian network for effective development of conversational agent," in *Intelligent Data Engineering and Automated Learning*, vol. 2690. New York: Springer-Verlag, 2003, pp. 1–9.

[21] G. Ferguson, J. Allen, and B. Miller, "TRAINS-95: Towards a mixed-initiative planning assistant," in *Proc. 3rd Int. Conf. Artif. Intell. Planning Syst.*, 1996, pp. 70–77.

[22] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The lumiere project: Bayesian user modeling for inferring the goals and needs of software users," in *Proc. 14th Conf. Uncertainty Artif. Intell.*, 1998, pp. 256–265.

[23] F. Bechet, J. Wright, A. Gorin, and D. Hakkani-Tür, "Detecting and extracting named entities from spontaneous speech in a mixed-initiative spoken dialogue context: How may I help you?" *Speech Commun.*, vol. 42, no. 2, pp. 207–225, 2004.

[24] S.-I. Lee, C. Sung, and S.-B. Cho, "An effective conversational agent with user modeling based on Bayesian network," in *Proc. 1st Asia-Pacific Conf. Web Intell.: Res. Develop., Lecture Notes in Computer Science,* vol. 2198. Berlin, Germany: Springer-Verlag, 2001, pp. 428–432.