Cognitive Model of Speech Dialogue

Yuichiro Anzai and Kaoru Hiramatsu
Department of Computer Science, Keio University
3-14-1, Hiyoshi, Kouhoku-ku, Yokohama 223 JAPAN
e-mail: anzai@aa.cs.keio.ac.jp, hiramatu@aa.cs.keio.ac.jp

1. Introduction

A speech dialogue is advantageous over other ways of communication between human and computer in that it needs no special training. While there have been a number of research projects on speech dialogue system, those systems are generally not able to work in dynamical environments. It is principally because they lack consideration of non-verbal information, and also relation between verbal and non-verbal information. Actually, human conversation contains, or is affected by, much non-verbal information that is somehow shared by speakers and listener.

"Mental models", often referred particularly in the field of cognitive science, may include this aspect. That is, integration of verbal and non-verbal information that represent the utterance situation for participants in conversation.

We believe that most of the speech dialogue systems for real-world application must somehow include formulation and implementation of mental models in the above sense. Thus our goal has been to model and implement a dynamical construction process of mental models under the domain of human-robot speech dialogue communication.

2. Design Principles

In order to approach our goal, we designed and implemented three speech dialogue systems with consideration of non-verbal information on autonomous mobile robot named ASPIRE[1]. ASPIRE has various kinds of sensors such as ultrasonic sensors, touch sensors, etc., which are controlled by multiprocessors. These speech dialogue systems are executed on PULSER[2], the operating system for ASPIRE developed in our laboratory. In this section, we describe three systems in detail.

3. Utterance Generation with Attention Mechanism

Linta-II[3] is an utterance generator which is implemented on the autonomous mobile robot ASPIRE for flexible human-robot interaction. Using what we call Attention Mechanism, Linta-II can generate situated utterances with external world information obtained from the robot's sensors. The Attention Mechanism consists of entities that make simple symbolic manipulation, and pay involuntary and voluntary attention.

Involuntary attention acquires an external world event immediately when it happens. Voluntary attention is activated by an action control unit of the robot, and makes symbolic manipulation in top-down fashion. With these two mechanisms, Linta-II can generate situated utterances without declarative constraints. Linta-II was implemented using what we call 'direct interrupt mechanism' offered by PULSER.

It consists of the attention mechanism, the utterance module and the action executive.

The application modules which are the utterance module and the action executive in the figure 1, use sensory data from the attention mechanism.

Figure 2 shows Attention Mechanism on Linta-II. Attention Mechanism provides sensor information which is used by application modules.

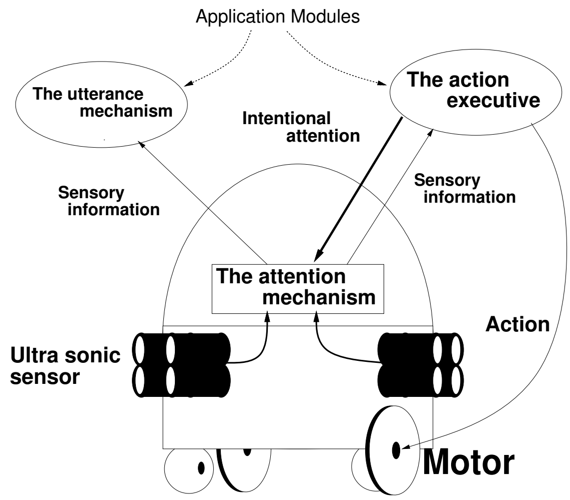


Figure 1:Architecture of Linta-II.

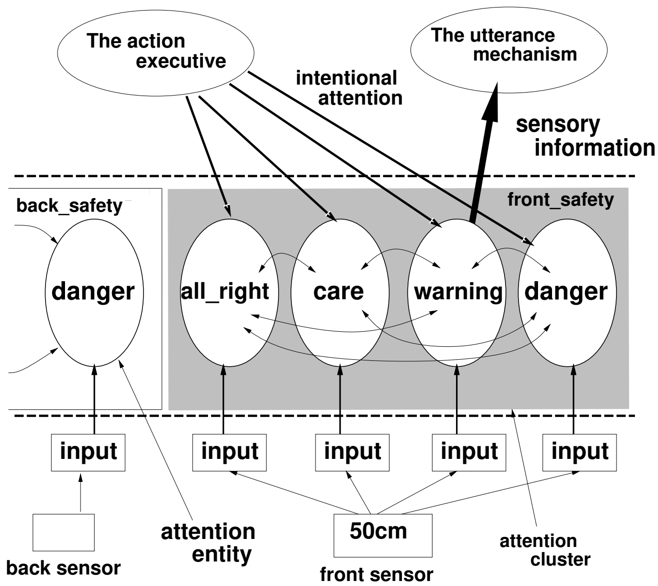


Figure 2: Outline of Linta-II's attention mechanism.

Linta-II was evaluated according to the number of declarative constraints used in two types of utterance generation: the one is with attention mechanism and the other is without attention mechanism. This result suggests that the attention mechanism is effective to reduce the number of the constraints.

4. Reactive Dialogue Planner with Motivation Mechanism

One of the weaknesses of Linta-II is that no "motivation mechanism" is implemented. Also, the system, though it is reactive to changes of external environments, needs more capacity for reactiveness to avoid danger. Reactive Dialogue Planner was designed to motivate utterances by itself, and also can manage reactiveness, by incorporating a reactive planning algorithm like Georgeff's PRS (Procedural Reasoning System)[4].

Our Reactive Dialogue Planner is based on the assumption that scenarios are limited in which a speech dialogue with a autonomous mobile robot is embedded. Dialogue plans are described as reactive plans that possess an ability to respond immediately to changes of linguistic context obtained from speech inputs and changes of external environments obtained from sensory inputs.

When the situation matches more than one plan, the dialogue manager mediates these plans, based on some heuristic strategies. When there is no plan being executed, the most urgent action or plan for the situation is selected. In the other cases the dialogue manager mediates matched plans in plans being executed: if no plan matches to the situation, the action or plan is selected in the aforementioned way.

It is assumed that plans have priority values with the range between zero and one. The priority value of the plan being executed is reset to one when the dialogue of the plan is selected. When the plan called from another plan is selected, the priority value of the plan calling the selected plan is updated in some predetermined fashion. We designed and implemented this speech dialogue system that includes the reactive planning subsystem to realize utterance motivation and more reactiveness.

To accomplish that, we first implemented a voice recognition board and a task planner on the ASPIRE. Then, the dialogue planner mentioned above was implemented, which completed our system for speech dialogue with autonomous mobile robots.

The overall system integrates goal-oriented dialogue planning and reactive dialogue planning. The former is in a sense similar to those in speech dialogue systems proposed in the conventional work on human-computer interaction. The latter is concerned with external and contextual changes. This integration made our speech dialogue system enable to realize human dialogue with autonomous mobile robots, whose "mental models" are reconstructed according to dynamical changes in external environments.

5. User Identification System and Sensor Manager

The above systems are reactive to changes of external environment, but the utterance continuity was not considered. The processing speed of the voice generator is much slower than the change speed of external environment. If the user moves restlessly, the robot cannot make effective utterances because of the changes of external environment. However, if the robot system react on changes of external environment immediately, it will make fragmentary utterances to the user. According to the difference of the speed, it is difficult to make flexible dialogue.

On our new system, we import the device hold time for the action continuity. To adjust the hold time, we can arrange the timing that the robot switches the user to talk to. The

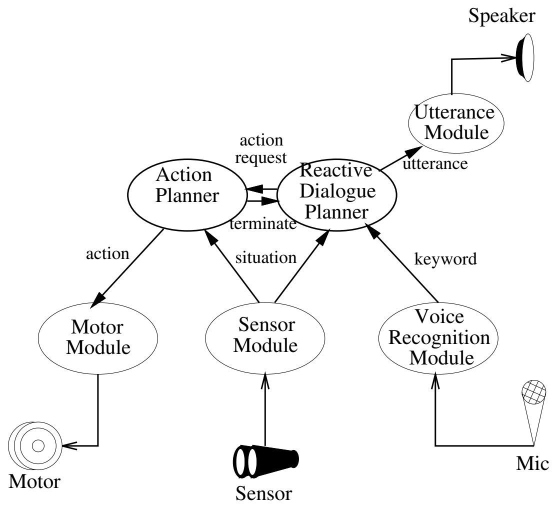


Figure 3: Outline of Reactive Dialogue Planner.

timing is settled by an user of the robot or application modules on the robot.

To implement this, we designed two modules, an user identification system and Sensor Manager.

The user identification system[5] consists of Identification Pendant and receivers on the robot. Identification Pendant is put on by an user of robot and transmits user's ID by infrared signals. The receivers get signals and then the robot identify the user near it. This system provides non-verbal information, such as user identification, user's direction, etc. Figure 4 is a overview of this identification system.

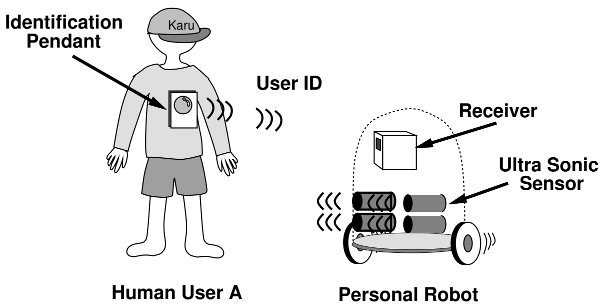


Figure 4: Overview of User Identification System.

To integrate information of external environment, Sensor Manager is designed to manage several kinds of sensory inputs concurrently. Figure 5 shows outline of the system. Inputs of manager are ultra sonic sensor, voice recognition system, optical-fiber gyroscope, etc. Sensor Manager has action rules. The action rule form is (sensory input)-¿(primitive action). If the sensory inputs match condition of rules, an primitive action starts to control output devices.

In the environment that changes restlessly, many action rules match external environment

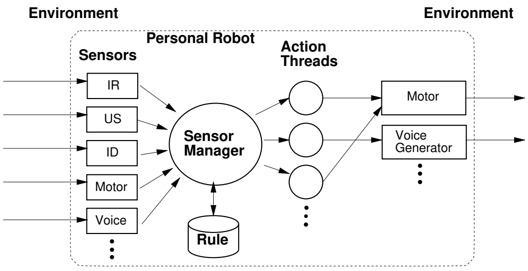
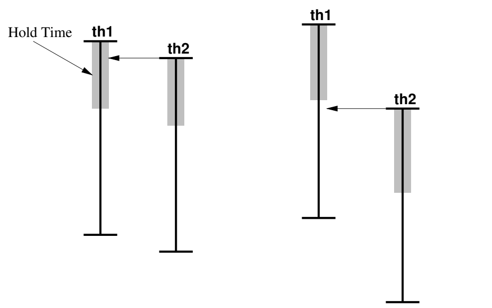


Figure 5: Outline of Sensor Manager.

and many action threads will be created. To avoid conflicts among action threads using same output device, we employ thread's priority. The thread's priority is arranged by, 1)emergency level of the thread, 2)hold time of output device and 3)class of sensory device.

The hold time is a parameter of reactivity and important for the dialogue system on robots to hold generating voice utterance. If two action threads conflict on output device, priority of threads is settled by hold time of previous action thread. While the hold time of the action thread, the action thread created previously is able to hold a output device (Figure: 6(a)). After the hold time, the thread created later is able to hold (Figure: 6(b)).



(a)th2 is created while hold time of th1. (b)th2 is created after hold time of th1.

Figure 6:Hold time of threads.

If the hold time settle utterance time, the robot makes the utterance completely but cannot react to the changes of external environment. On the other hand, if the robot conducts itself only based on the external environment change, the robot reacts as soon as the external environment change. However, the robot will switch the focusing user too often to makes the fragmentary utterance.

The hold time is important for physical action of robots, especially in speech dialogue system, to make flexible human-robot interaction. But many factors are related with the hold time. In the case of human-human communication, the hold time may be settled by the changes of external environment, sentence of utterance, social relationship, etc.

In this system, we only offer the structure to arrange the hold time. The hold time is able to settle for experiment or able to control from the upper module such as dialogue planner.

6.Conclusion

We believe that most of the speech dialogue systems for real-world application must somehow include formulation and implementation of mental models. To model and implement a dynamical construction process of mental models, we used the domain of human-robot speech dialogue communication and implemented three modules for a dialogue system, Linta-II, Reactive Dialogue Planner and Sensor Manager. According to these modules, the dialogue system can get non-verbal information and changes of the external environment and apply to dialogue systems.

To make more flexible dialogue systems, we have to handle other abstract factors such as social relations.

References

[1] N.Yamasaki and Y.Anzai."The Design of the Personal Robots". In Proc. of JSME Annual Conference on Robotics and Mechatronics, pages 51–56, 6 1992. (In Japanese).

[2] T.Yakoh, T.Sugawara, T.Akiba, T.Iwasawa and Y.Anzai. "PULSER: A Sensitive Operating System for Open and Distributed Human-Robot-Computer Interactive Systems". In Proc. IEEE Int. Workshop on Robot and Human Communication, pages 404–409, Tokyo, 9 1992.

[3] M.Imai, K.Hiraki and Y.Anzai. "Utterance Generator with Attention mechanism". The Transactions of The Institute of Electronics, Information and Communication Engineers D-II, J-77-D-II(8):1447–1456, 1994. (In Japanese).

[4] M.P.Georgeff and F.F.Ingrand. "Decision-making in an embedded reasoning system". In IJCAI-89, pages 972–978, 1989.

[5] K.Hiramatsu and Y.Anzai. "User Identification in Human Robot Interaction Using Identification Pendant". In Proc. of the 6th International Conference on Human-Computer Interaction, pages 20B–237–242, 1995.

Principal Publications

[1] M.Imai, K.Hiraki and Y.Anzai, "Utterance Generator with Attention mechanism", The Transactions of The Institute of Electronics, Information and Communication Engineers D-II, Vol.J77-D-II, No.8, 1994. (In Japanese)

[2] Y.Uyama, "The Design and Implementation of Reactive Spoken Dialogue System", Master Thesis, Keio Univ, 1995. (In Japanese)

[3] K.Hiramatsu and Y.Anzai, "User Identification in Human Robot Interaction Using Identification Pendant". In Proc. of the 6th International Conference on Human-Computer Interaction, pp.20A-237-242, 1995.

keyword: attention, utterance, human-robot interaction, reactivity, user identification