Paper:

Commitment-based Natural Language Interface System for Robots

Yasushi Nakauchi*,Takeshi Takahashi, Piyawat Naphattalung*,Takashi Matsubara* and Eiichi Kashiwagi***

*Department of Computer Science, National Defense Academy
1-10-20, Hashirimizu, Yokosuka, 239-8686 Japan
E-mail: {nakauchi,g40048,matubara,kashi}@nda.ac.jp
**Officer Candidate School, Japan Ground Self Defense Force
2728 Koura-uchi-machi,Kurume, 839-0852 Japan
E-mail: trbt1979@k3.dion.ne.jp
[Received June 15, 2002; accepted July 24, 2002]

In this paper, we first discuss how the commitment in human-robot natural language conversation is important and the lack of the commitment leads to breakdowns in conversation. Then, we propose a human-robot conversation model, which takes the breakdown into consideration. We implemented a natural language interface system of a mobile robot, based on the proposed conversation model. Then, we conducted experiments with subjects and confirmed that the proposed interface system provides more natural conversation between human and robot with fewer breakdowns.

Keywords: conversation theory, commitment, breakdown, human-robot interaction

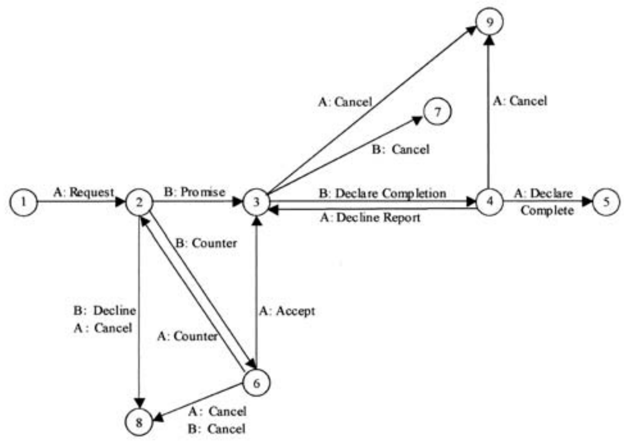

Fig. 1. Converation theory model for human.

1. Introduction

Recent progress in robotics technologies is making it promising to use robots in everyday life. These robots will be used by ordinary people at ordinary places (i.e. at home, offices, and hospitals). For ordinary people, a natural spoken language interface is considered the most desirable, since they could interact with robots as they do to other people.

There are many robot systems, which accept orders by natural spoken language[4,12]. It is reported that simple commands such as "move forward" are misunderstood at 10% by using IBM ViaVoice[10]. If a robot fails to recognize a command issued by a user, the user has to cancel it and reissue the command. Even if a robot recognizes the order correctly and tries to execute the command, since the robot has to perform in the real world, the robot may fail to execute it. Then, the user has to correct the incorrect behavior by voice. These phenomena lead to **breakdowns** [14] in the conversation.

We propose a human-robot conversation model, which takes breakdown into consideration. We also implement a natural language human-robot conversation system by using a mobile robot and confirm its efficiency by experiments.

2. Natural Language Interface Systems

2.1. Human-to-Human Conversation Model

In the context of human-to-human dialogue, Winograd and Flores pointed out that breakdowns occur due to the lack of commitments between humans[14].

The commitment they define is the state where participants in dialogue owe certain responsibilities and one expects the responsibility to be borne by the others. The breakdown is the state where participants recognize something wrong in the conversation and it is triggered by the lack of commitments.

They also proposed a human-to-human conversation model, which anticipates breakdowns (see **Fig.1**).

In the figure, the arcs represent actions and the nodes represent states in conversation. The arc **request** means to place an order (task), **promise** means to accept the

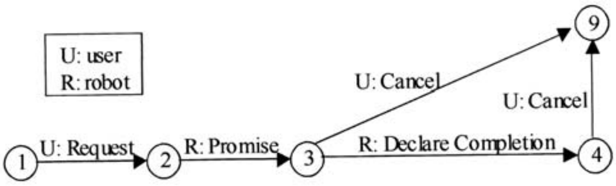

Fig. 2. Conventional human-robot conversation model.

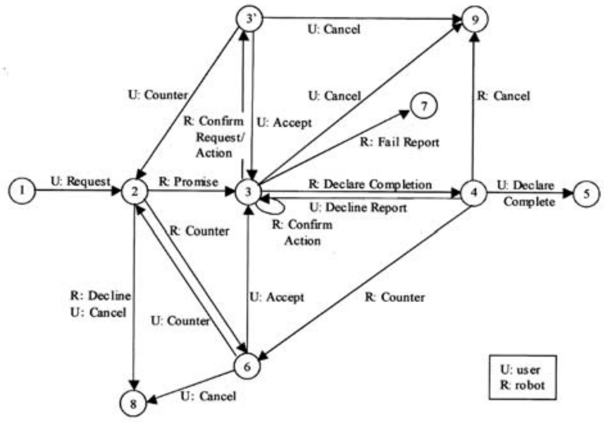

Fig. 3. Commitment based human-robot conversation model.

request or counter offer, **declare completion** means to declare the completion of the promised task, **declare complete** means to declare the completion of the task, **decline report** means to reject the report telling the completion of the task and making to carry the task out again, **accept** means to accept and promise the proposed counter offer, and **cancel** means to discontinue the task or request.

A typical course of conversation between user **A** and **B** is as follows.

User **A** starts the conversation by making request to user **B**. In the state 2, there are five possibilities. User **B** could accept, reject or negotiate a change on initial conditions, or user **A** could cancel the request or negotiate a change on initial conditions. Following the normal course of the conversation, at the end (state 5, 6, 8 and 9), the actors will have a happy end without breakdowns.

2.2. Breakdowns in Human-Robot Conversation

Especially for human-robot conversation, we presume the following four kinds of breakdowns, which are not included in the human-to-human conversation model shown in **Fig.1**.

(1) Recognition error breakdowns

If the error in voice recognition occurs, the robot may execute wrong order. (i.e. if the robot recognized as "turn **right**" though a user said "turn **left**", the robot will turn to the wrong direction). This kind of breakdowns force the user to cancel the execution and reissue the correct order.

(2) Motion error breakdowns

Even if the robot tries to execute a correct order, since the robot has to perform in the real world, it may not execute the task correctly. This kind of breakdowns force the user to reissue the command for correcting the robot performance.

(3) Externalization error breakdowns

It is not always obvious for a user if a robot is performing the ordered task or not. Some task executions may not accompany any conversation or gesture, so the user may issue the next command while the robot is still performing and it will confuse the robot. This kind of breakdowns are caused by the lack of externalization of the robot internal states.

(4) Motion continuity breakdowns

In general, sequential orders from a user has a meaning as a context.

For example, after a user ordered a robot to bring a medicine, the user will succeedingly order a glass of water.

These sequential orders are considered as a routine work and if the user has to ask these orders one by one and every time, it will be stressful for the user.

To the contrary, if a robot offers these desirable orders to the user, it will prevent these breakdowns.

2.3. Human-Robot Interface

So far, there are already several speech command acceptable robots in the market[4,12]. Most conventional robots are based on a simple dialogue model, which just recognize a voice order and perform it (see **Fig. 2**). And the user repeats the simple order and execution cycle until a task is completed. Since voice recognition systems contain a certain percentage of recognition errors [10], these simple methods lead to recognition error breakdowns.

In such breakdowns, efforts to resolve ambiguities in spoken language are being made. The conversation system Linta resolves ambiguity by using sensor information observed by a robot[13]. Jijo-2 and Robovie robots resolve ambiguity by using situational context in conversation[8,5].

In externalization error breakdowns, CERO is resolving by regularly reporting internal states of a robot to the user[3].

Though the above systems are resolving each aspects of breakdowns in human-robot conversation discussed in section [3], they are not resolving commitments nor breakdowns as a whole. The conversation model proposed by Winograd and Flores has applied to groupware (human-to-human interface system), The Coordinator[1], but there is no conversation system, which considers the above breakdowns in human-robot interaction.

3. Commitment-based Human-Robot Conversation Model

To overcome the four kinds of breakdowns discussed in section 2.2, we modified the conversation model proposed by Winograd and Flores as follows (see **Fig. 3**).

First, for (1) recognition error breakdowns, we added node 3'. We have drawn the arc from node 3 to 3', so a robot confirms the recognized command. At the state 3', a user has three choices: accept the correctly recognized command (accept arc, which will be automatically followed in a certain time period has passed even if a user explicitly confirm it by voice (as the implicit agreements)), correct the misunderstood command (counter arc), or cancel the command (cancel arc).

For (2) motion error breakdowns, we made a robot to also follow the arc from node 3 to 3' when it performs a certain task for a long time. By this speech act, a robot could confirm to the user if it is performing the task correctly. For the (3) externalization error breakdowns, we have added the looped arc on node 3, so a robot explicitly expresses the internal states to a user periodically. Finally, for (4) motion continuity breakdowns, we added the arc from node 4 to 6, so a robot could propose the presumable order as a counter offer.

With these modifications, conversations of a robot and a user will be improved as follows.

($_iX_j$ indicates the transition from state i to j by actor X(U: user or R: robot)'s speech act.)

(1) Recognition error breakdowns

Conversational model (see **Fig. 2**):
$_1U_2$: "Turn to the **right**."
$_2R_3$: "I will turn to the **left**."
(Recognition error occurred and the robot turns to the **left**.)
$_3U_9$: "Stop it."
$_1U_2$: "Turn to the **right**."
$_2R_3$: "I will turn to the **right**."
$_3R_4$: "I have completed the task."
Proposed model (see **Fig. 3**):
$_1U_2$: "Turn to the **right**."
$_2R_3$: "All right,
$_3R_{3'}$: I will turn to the **left**."
$_{3'}U_2$: "**Right** direction."
$_2R_3$: "All right,
$_3R_{3'}$: I will turn to the **right**."
$_{3'}R_3$: "Good."
$_3R_4$: "I have completed the task."
$_4R_5$: "Thank you."

With the conventional model, when a voice recognition error occurred, the user has to terminate the wrong movement and reissue the correct one again.

With the proposed model, since the robot asks the user for confirmation, the user could know the voice recognition error and correct by saying only misunderstood part (in this case the direction to turn).

(2) Motion error breakdowns

When a robot is moving straight forward, the direction has shifted to the left.

Conventional model:
$_3R_9$: "Stop."
$_1U_2$: "Move a little to the right."
$_2R_3$: "I will move a little to the right."
Proposed model:
$_3R_{3'}$: "**Is this the right direction?**"
$_{3'}U_2$: "No."
$_2R_6$: "**Which direction should I move?**"
$_6U_2$: "Turn a little to the right."
$_2R_3$: "All right,
$_3R_{3'}$: "I will move forward to the right."

With the conventional model, the user has to monitor the robot movement. When the user found the malfunctioning, the user has to stop the robot once, and then reissue the new command.

With the proposed model, the robot confirms if it is still executing correctly, so the user could correct the wrong part in the robot movements and could keep the robot working as the user intended.

(3) Externalization error breakdowns

Conventional model:
$_1U_2$: "What is that (in front of you)?"
$_2R_3$: "I will investigate it."
(Robot is investigating by using sensors)
$_1U_2$: "Move backward a little."
Proposed model:
$_1U_2$: "What is that?"
$_2R_3$: "All right."
$_3R_{3'}$: "I will investigate it."
$_{3'}U_3$: "Good."
$_3R_3$: "**Investigating.**" (periodically)
$_3R_4$: "This is a film case."
$_4U_5$: "Thank you."

With the conventional model, since the user could not know if the robot is still investigating, the user considered the robot is stacked and issued the next command.

With the proposed model, since the robot explicitly explains the internal states via voice, the user could wait for the robot performance confidently.

(4) Motion continuity breakdowns

Conventional model:
$_1U_2$: "What is that (in front of you)?"
$_2R_3$: "I will investigate it."
$_3U_4$: "This is a film case."
$_1U_2$: "Lower the arm."
$_2R_3$: "I will lower my arm."
$_3R_4$: "I have lowered my arm."

Table 1. Typical phrases which Khepera accepts.

Item	Phrase
Mobile base	"Turn to the right/left."
	"Move forward/backward."
	"Move to the right/left." (The robot turns right/left then moves forward)
Arm	"Lower/Raise the arm."
Gripper	"Close/Open the gripper."
Sensors	"What is it?" (The robot investigates the object in front of the robot)
Adverbs	"a little", "more/much"
Ack	"Thank you."

Fig. 4. Mobile robot Khepera.

${}_1U_2$: "Close the gripper."
${}_2R_3$: "I will close my gripper."
${}_3R_4$: "I have closed my gripper."

Proposed model:

${}_1U_2$: "What is that?"
${}_2R_3$: "All right,
${}_3R_{3'}$: "I will investigate it."
${}_{3'}R_3$: "Good."
${}_3R_3$: "Investigating." (periodically)
${}_3R_4$: "This is a film case."
${}_4R_6$: **"Shall I lower my arm?"**
${}_6R_3$: "Good."
${}_3R_4$: "I have lowered my arm."
${}_4R_6$: **"Shall I close my gripper?"**
${}_6U_3$: "Good."
${}_3R_4$: "I have closed my gripper."
${}_4U_5$: "Thank you."

With the conventional model, the user has to issue the sequence of orders for grasping a film case one by one.

With the proposed model, the robot refers the command history issued by the user, so the robot can expect the sequence of orders for grasping after it found a film case. To the contrary, the robot asks permission for executing the sequence of orders. This will realize more natural and pleasant conversation than conventional ones.

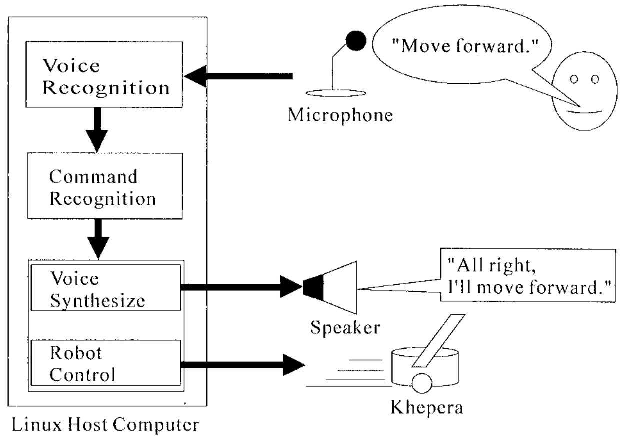

Fig. 5. Schematic diagram of the dialogue system.

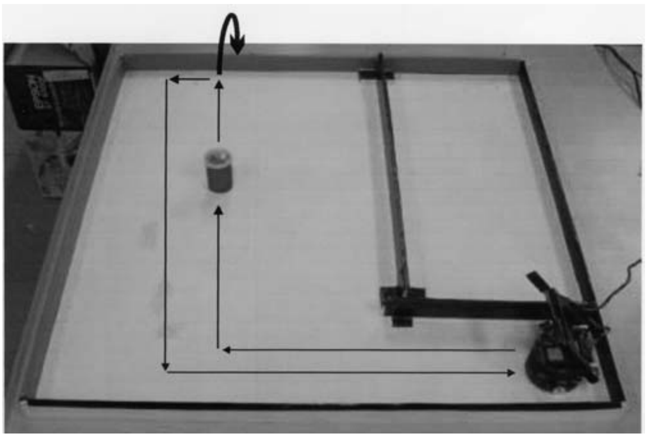

Fig. 6. Experimental environment.

4. Implementation

We implemented the human-robot dialogue system based on the proposed conversation model on a mobile robot, Khepera (see **Fig. 4).** Khepera only has four degree of freedom (i.e., two for mobile base which enables to translate and to turn and two for gripper turret which enables to flip up/down the arm and to open/close the gripper). Khepera only has eight infrared very short - range sensors. The control flow of the dialogue is as shown in **Fig. 5**. We employed the speech recognition system Julius[7] and the speech synthesize system developed by Create System Development Co., LTD.

Since sensors and actuators Khepera has are rather restricted, we implemented all possible dialogue patterns for speech. Typical phrases which Khepera accepts are as shown in **Table 1**. We also implemented alternative phrases such as "grasp it" for "close the gripper" to increase voice recognition quality.

Examples of actual conversations are as shown in sec-

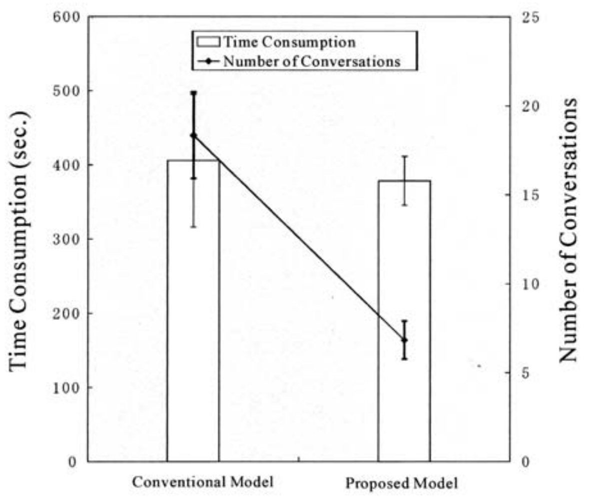

Fig. 7. Consumption time and number of conversations per task.

tion 3.

As for the counter offers the robot proposes, we recorded the history of commands a user issued, and when a pair of the succeeding two commands has been observed as it has a correlation (exceeded a certain threshold), we made the robot propose it[9]. All these systems are implemented on a Linux host computer (Pentium IV, 2GHz) and Khepera is controlled via an RS-232c serial communication network (see **Fig. 5**).

5. Experiments and Discussion

To confirm the efficiency of the proposed system, we conducted experiments with subjects. The experimental environment for the robot is as shown in **Fig. 6**. The size of the field is 60cm × 60cm × 2.5cm. The task that the subjects perform was to move up the robot in front of the film case, to make the robot to recognize the object[1] (if it is a film case or the wall), to grasp it, to move in front of the wall, to put it outside of the field, and to move back to the initial position (see the arrows on **Fig.6**). Subjects were five undergraduate students in the computer science department.

We have prepared two types of natural language interface. One is the conventional dialogue, based on the simple conversation model in **Fig. 2**. The other is the proposed dialogue, based on the conversation model in Fig. 3. To eliminate experiencing side-effects (subjects will be accustomed with the task while they are performing experiments), we have asked each subject to perform the task five times for each dialogue system with the random order. The averaged consumption time and number of conversations for each trial are as shown in **Fig.7**. The number of conversation means the number of series of conversation in each conversation model (i.e. In the case of **Fig. 3**, the conversation which starts from node 1 and finishes on node 5, 7, 8, or 9 is counted as one conversation.).

From the **Fig.7**, with the proposed dialogue system, we observed that the consumption time per task has improved a little and the number of conversations is decreased from 18 to 7.

We also videotaped subject performance and counted the number of voice recognition errors and the number of motion errors. The averaged number of voice recognition errors per task was 1.3 and the averaged number of robot motion errors per task was 2.0. Below, we discuss how the proposed system worked for preventing breakdowns.

(1) Recognition error breakdowns

Voice recognition errors were observed about 1.3 times per task.

With the conventional model, when these errors occurred, a user stopped the wrong movement of the robot and then issued the command for the cancellation (i.e., when the robot misunderstood the phrase as “**move** to the right” in stead of “**turn** to the right”, the robot turned right and then moved forward, so the user had to stop the motion and then had to move the robot backward to cancel the movement.)

With the proposed model, since a user could know the error before the robot executes it, the user corrected the misunderstood word (at the node 3’ in **Fig. 3**), so recognition error breakdowns were not observed.

(2) Motion error breakdowns

Motion errors were observed about 2.0 times per task.

Most cases where error occurred were as follows.

When a user turned the robot towards the film case and made the robot move straight forward (towards the film case), the proceeding direction curved gradually and the robot could not face the film case.

With the conventional model, a user had to stop the robot when the robot direction curved, and then correct the direction by turning the robot.

With the proposed model, the robot asked if the translating direction is still correct when it was translating, so the user could correct the direction by saying such as “a little right” (the counter offer shown as the arc from the node 3’ to the node 2 in **Fig. 3**), so motion error breakdowns were not observed.

(3) Externalization error breakdowns

With the conventional model, when the robot is investigating the object in front of it silently, a user sometimes issued the next command, so the robot could not estimate the direction face to the object. With the proposed

1 When the robot recognized the object in front of it as a film case, the robot rotates so the f ilm case may become the front.

model, since the robot periodically declares that it is still investigating, externalization error breakdowns were not observed.

(4) Motion continuity breakdowns

With the proposed model, when the robot came close to the wall with the film case in the gripper, the robot itself inferred the presumable next user command from the past command history with the current situation, and then proposed a user for releasing the file case on the out side of the field. Also, when the robot detected a film case, the robot proposed a user for catching it. In these cases, what a user has to do was just permitting these proposals. With these incidents, it has confirmed that a user could smoothly operate the robot. This phenomenon is also observed as the fewer number of conversations by the proposed model as shown in **Fig.7**.

6. Conclusion

We proposed a conversation model for human-robot natural language conversations, focusing on breakdowns in human-robot conversation, and proposed a commitment-based conversation model. In experiments, we confirmed that the proposed system prevents breakdowns and provides smooth conversation between the human and robot.

The conversation system we developed was rather simple because of the maneuvering and sensing ability Khepera has, so we plan to extend sensing ability by employing a CCD camera on Khepera. For externalization of the robot status, we are planning to implement gestures as well as voice.

In this paper, we assumed that the task for Khepera to perform was to handle film cases. In other directions for expanding our work, we plan to apply our conversation model to applications such as secretary robots[8] and service robots[6,11].

References:

1) F. Flores, et al., "Computer Systems and the Designs of Organizational Interaction", ACM Trans.OIS, 1988.
2) M. Fujita, et al., "An Autonomous Robot that Eats Information via Interaction with Humans and Environments", In Proc. of IEEE Itnl. Workshop on Robot and Human Interactive Communication, pp.383–389, 2001.
3) A. Green and K. Severinson-Eklundh, "Task-oriented Dialogue for CERO: a User-centered Approach", In Proc. of IEEE Itnl. Workshop on Robot and Human Interactive Communication, pp.146–151, 2001.
4) http://www.sok.co.jp/r d/robo.html
5) H. Ishiguro et al., "Robovie: A robot generates episode chains in our daily life", In Proc. of Itnl. Symposium on Robotics, pp.1356–1361, 2001.
6) Y. Iwakura, Y. Shiraishi, Y. Nakauchi and Y. Anzai,"Multi-agent Interface Architecture for Real-World Oriented Distributed Human Interface Systems", In Proc. of IEEE SMC'97, pp.4115–4120, 1997.
7) A. Lee, T. Kawahara, and K. Shikano, "Julius–an open source real-time large vocabulary recognition engine", In Proc. of European Conf. on Speech Communication and Technology, pp.1691–1694, 2001.
8) T. Matsui et al., "Integrated Natural Spoken Dialogue System of Jijo-2 Mobile Robot for O.ce Services", In Proc. of AAAI, 1999.
9) T. Mitchell, "Machine Learning", McGraw-Hill,1997.
10) H. Mizoguchi, T. Okabe, K. Hiraoka, T. Shigehara and T. Mishima, "Voice Interface of a Mobile Robot for Human-Robot Interaction Research", Proc. of JSME Robomech'00, 1P1-77-121, 2000.
11) Y. Nakauchi and R. Simmons, "A Service Robot for Peopled Environments –Finding Lined People by Stereo Vision and Standing in Line–", Prod. of 2001 IEEE/ASME Itnl. Conf. on Advanced Intelligent Mechatronics, pp.1093–1098, 2001.
12) H. Nishizawa, et al., "The Hardware Architecture of Personal Robot PaPeRo", In Proc. of RSJ2001, pp.277–278, 2001 (in Japanese).
13) M. Satoh, K. Hiraki and Y. Anzai, "A Dialogue with a Robot: A Design and Implementation of Speech Dialogue System using Sensor Information", In. Proc. of JSAI SIG-SLUD-9202-3, pp.410-417, 1992.
14) T. Winograd and F. Flores, "Understanding Computer and Cognition", Addison-Wesley, 1986.

Name:
Yasushi Nakauchi

Affiliation:
Associate Professor, Dept. of Computer Science, National Defense Academy

Address:
1-10-20, Hashirimizu, Yokosuka, 239-8686 Japan

Brief Biographical History:

1993- Research Associate at National Defense Academy
1994- Assistant Professor at National Defense Academy
1998-1999 Visiting Scientist at Carnegie Mellon University
1998- Associate Professor at National Defense Academy

Main Works:

- "A Social Robot that Stands in Line", Autonomous Robots, Vol.12, pp.313-324 ,2002.

Membership in Learned Societies:

- The Japan Society of Mechanical Engineers (JSME)
- The Robotics Society of Japan (RSJ)
- The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Takeshi Takahashi

Affiliation:
Sergeant Major at Officer Candidate School, Japan Ground Self Defense Force

Address:
5601-2, Teradoko, Matsuo, Iida, Nagano, 395-0822 Japan

Brief Biographical History:
1998- Dept. of Comp. Sci., National Defense Academy
2002- Sergeant Major at Officer Candidate School, JGSDF

Main Works:
- "A Natural Language Interface System for Mobile Robots based on Conversation Theory", Proc. of JSME Robomec'02, 2P2-L03, 2002.

Name:
Takashi Matsubara

Affiliation:
Research Associate, Department of Computer Science, National Defense Academy

Address:
1-10-20, Hashirimizu, Yokosuka, 239-8686 Japan

Brief Biographical History:
1991- Reseach Associate at National Defense Academy

Main Works:
- "On M-Sequence Generators with Fault Detecting Checkers and Their Characteristics",IEICE Trans. D-I, Vol.J81-D-I, No.3, pp.283-291, 1998.

Membership in Learned Societies:
- The Institute of Electronics, Information and Communication Engineers (IEICE)
- The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Piyawat Naphattalung

Affiliation:
Master Course Student, School of Computer Science, National Defense Academy

Address:
1-10-20 Hashirimizu, Yokosuka, 239-8686 Japan

Brief Biographical History:
1993-Armed Forces Academies Preparatory School,Thailand
1995-Chulachomkloa Royal Military Academy,Thailand
1996-Dept. of Comp. Sci., National Defense Academy
2001-Master Course Student at School of Comp. Sci., National Defense Academy

Main Works:
- "A Natural Language Interface System For Mobile Robots based on Conversation Theory",Proc. of JSME Robomec'02, 2P2-L03, 2002.

Membership in Learned Societies:
- The Japan Society of Mechanical Engineers (JSME)
- The Robotics Society of Japan (RSJ)

Name:
Eiichi Kashiwagi

Affiliation:
Professor, Department of Computer Science, National Defense Academy

Address:
1-10-20, Hashirimizu, Yokosuka, 239-8686 Japan

Brief Biographical History:
1966- Research Associate at National Defense Academy
1969- Assistant Professor at National Defense Academy
1974- Associate Professor at National Defense Academy
1989- Professor at National Defense Academy

Main Works:
- "Proposal of Novel Eight-Figure Kernel of Time-Frequency Analysis",IEICE Trans. A, Vol.J81-A, No.6, pp.907-915, 1998.

Membership in Learned Societies:
- The Institute of Electronics, Information and Communication Engineers (IEICE)
- Information Processing Society of Japan (IPSJ)
- The Society of Instrument and Control Engineers (SICE)