Towards Robot as an Embodied Knowledge Medium — Having a robot talk to humans using nonverbal communication means

Taku Ohya
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
ohya@ii.ist.i.kyoto-u.ac.jp

Tatsuya Hiramatsu
Undergraduate School of Information and Mathematical Science, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
hiramatsu@ii.ist.i.kyoto-u.ac.jp

Yong XU
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
xuyong@ii.ist.i.kyoto-u.ac.jp

Yasuyuki Sumi
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
sumi@i.kyoto-u.ac.jp

Toyoaki Nishida
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
nishida@i.kyoto-u.ac.jp

Abstract

Recently, several types of interactive humanoid robots have been developed, and it has become necessary for humans to communicate with robots and treat them not only as intelligent devices but also as cooperative partners. In this paper, by using the task of assembling/disassembling a bicycle as an example, we propose a presenter robot system that can play tutorial videos with a display installed on its arm. In this manner, the robot can convey information regarding the steps and techniques of the operation, which is difficult to express verbally; thus, the robot can function like an operating manual. The robot moves to a suitable position and plays the appropriate video by utilizing the positional information of the user, the objects, and the robot. Since the robot can adjust the position of the display for easy viewing and change its position according to the user, the system can be adapted to different users. The implementation of this robot is expected to facilitate easy understanding by the user and enable him/her to perform a task efficiently.

Keywords: Presenter robot, display, Bayesian net.

1 INTRODUCTION

Several types of interactive humanoid robots have been developed in recent times. The increased level of human-robot interaction has made it necessary for humans to communicate with robots and treat them not only as intelligent tools but also as cooperative partners. A robot that is able to record audiovisual information can function as a recording medium, and this information can serve to enhance our knowledge. Therefore, it is important to consider robots as a medium for communicating knowledge among humans. Humanoid robots are different from software agents because they have a physical body. Therefore, it is possible for robots to communicate an individual's expertise through nonverbal methods.

Our group has already developed several types of robots, including a waiter robot system, a cleaning robot, and a listener robot, that have the ability to communicate with humans through nonverbal techniques. Our research objective is to develop a presenter robot system that will utilize information provided by a listener robot. The presenter robot can function as a knowledge medium and communicate an individual's expertise to other people.

In this paper, we describe the method of designing a presenter robot to function as a knowledge medium. We employ the assembling/disassembling of a bicycle as our research task and present the robot as a novel device for communicating knowledge: the robot plays tutorial videos with a display installed on its arm, thereby functioning as an alternative to standard operating manuals. The implementation of this robot is expected to facilitate the understanding of the user with regard to the operating steps and techniques, and thus enable the user to perform a task efficiently.

This paper is organized as follows. Section 2 explains the concepts of a social medium and the presenter robot. Section 3 describes the method of designing a presenter robot to perform a specific task. Section 4 presents our approach and the details of the implementation of the presenter robot. Section 5 concludes this study and outlines a direction for future work.

2 ROBOT AS A SOCIAL MEDIUM

2.1 ROBOT IN HUMAN SOCIETY

Most of the previous works in robotics have focused on establishing a model of the body movements of a robot. Tani [1] developed a robot that can learn and imitate human behavior. This robot utilizes a repetition algorithm implemented by a recurrent neural net that employs a parametric bias and divides behaviors into primitive behaviors by self organization. Since the robot does not require obvious teacher signals, its learning system is very robust. However, it does not focus on the ability necessary for carrying out a task in cooperation with a human user, namely, to understand the user's requirements and respond accordingly.

In order to provide support or collaborate with humans, a robot should ideally be capable of human-like movement. However, it is necessary for a robot to have the ability not only to move but also be sociable and communicate with humans.

With this consideration, Shiomi of Osaka University/ATR [2] attempted to create an experimental exhibition guide robot by using a humanoid conversation robot at a science pavilion. This robot acquires positional information of itself and the action history of a person from ubiquitous sensors installed inside the building and interacts with interpretive exhibits and people. The analysis of a subjectivity evaluation reveals that the interaction receives a favorable response from a person. However, a moderately high reputation value is obtained as a characteristic of the robot when it does not employ interaction and a quantitative evaluation is not performed.

Sidner et al. [3] focused on nonverbal communication between humans and robots and proposed a technique of using engagement behaviors such as nodding and gazing to attract the interest of humans; further, they experimented with a penguin-type robot that was implemented with these behaviors. The speech data and behaviors they used were limited and fixed. The experiments revealed a certain degree of influence on humans due to the engagement behaviors. However, neither did they focus on the acquisition of information from humans for generating new behaviors nor did they discuss how the technique might be adapted to different individuals.

Therefore, it is important to consider the robot not only as a partner but also as a medium that can communicate stored knowledge acquired though observation and learning.

2.2 HUMANOID ROBOT AS A SOCIAL MEDIUM

In order to consider the robot as a social medium that exchanges knowledge with humans (Figure 1), the robot needs to learn to interact with humans in a manner similar to natural human-to-human interactions. As is observed in human-to-human interactions, if the robot reacts to humans in an unnatural manner, they may sense that something is wrong, at worst, they may lose interest in the robot and stop interacting with it. In order to overcome this problem, the robot requires the ability to exchange knowledge with humans and accordingly, the ability to interact naturally with humans and learn appropriate forms of social interaction from them.

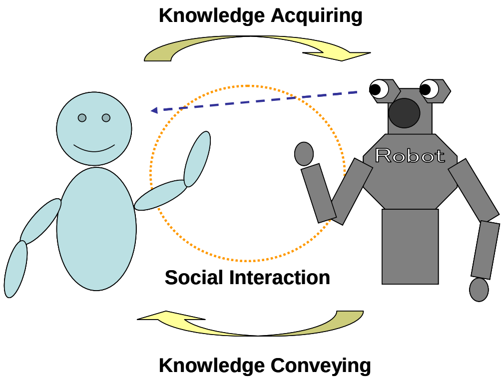

Figure 1 Humanoid Robot as Social Medium

2.3 LISTENER ROBOT

Ogasawara, a member of our group, has already developed a listener robot [4] that is able to acquire information from humans by listening. This robot has the ability to not only learn but also interact socially. It focuses on the utilization of joint attention and nonverbal communication to help humans feel comfortable. The listener robot can record as well as respond to a human speaker's explanation regarding the procedure of assembling a set of furniture or an electrical appliance. At present, Hiramatsu, another member of our group, is improving the listener robot to react to humans in a more sensitive manner and to create a more effective tutorial video by utilizing a zoom function.

2.4 PRESENTER ROBOT

The objective of our research is to develop a presenter robot that can efficiently communicate knowledge to humans, as shown in Figure 2. It achieves this objective through a natural style of interaction with a display that is integrated into its hand, which is designed to adapt to the user. We employ the task of assembling/disassembling a bicycle as an example, and the presenter robot presents the user (beginner) with video clips of another user (professional) performing the same task.

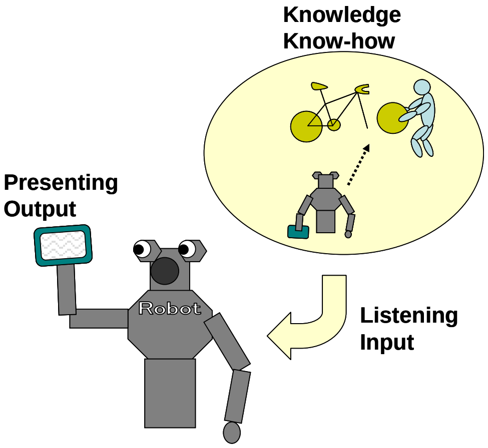

Figure 2 Listening and Presenting

A human mainly communicates knowledge by using two types of techniques: verbal (for example, speech/dialogue) and nonverbal (for example, gestures). The former makes it necessary for the robot to recognize and express information regarding the task explicitly through language; however, many practical difficulties are encountered in NLP (National Language Processing) technology. Our robot has few arm joints and does not have fingers; further, its capacity to express information is so poor that it cannot communicate expertise and steps and techniques of an operation to humans. Indeed, even among human beings, it is a difficult task to convey such information. One alternative is to use HMD (head-mounted display). However, with this system, interactions between the HMD user and the HMD are very limited; thus, it is difficult for a HMD user to communicate with others.

Our proposed method allows the user to easily understand the operational steps and techniques that are difficult to express using a manual. In the case of a fixed display, if the user changes his/her position he/she may not be able to view the display clearly. However, the display installed on the robot's hand can adapt to the changing positions of the user to simplify the task being performed by the user.

3 REQUIREMENTS FOR THE PRESENTER ROBOT

3.1 MAIN REQUIREMENTS

There are three main requirements for a presenter robot: The robot must be safe for humans, it should be able to communicate information to humans, and it should not inconvenience humans.

We demonstrate how these requirements are met by providing a solution to a specific task. In particular, we solve the important problem of developing a robot that is robust (resistant to noise) and stable (i.e., moves smoothly).

3.2 SUPPOSED TASK SCENARIO

In this paper, we employ the task of assembling/disassembling a bicycle as an example. The key operations are assembling/disassembling the front/rear wheel and opening/closing the front/rear brake. The steps involved in the assembling task can be summarized as follows. Firstly, the user assembles the rear wheel and closes the rear brake. Secondly, the user approaches the front of the bicycle and assembles the front wheel. Lastly, the user closes the front brake. In the next section, we present the details of the behaviors of the robot while following the steps listed above.

3.3 BEHAVIORS OF PRESENTER ROBOT

3.3.1 Approaching the Task Area

The presenter robot is required to shift to a suitable position depending on the user's current position, which is a very important requirement. If the robot is too close to the user, it may interfere with the operation being performed. On the other hand, if the robot positions itself at an excessively large distance from the user, it may become difficult to clearly view the display on the robot's arm. Moreover, the robot needs to shift its position to adapt to the changing positions of the user. However, abrupt motions of the robot would make it difficult for the user to view the display. Accordingly, the movements of the robot require to be stabilized for some situations.

3.3.2 Adjusting Display for Easy Viewing

Adjusting the position of the display to be in proximity to the user's work area is the most effective method of communicating information. Therefore, it is also important for the robot to control its arm movements such that the display can be viewed clearly. If the task area or task operation changes, the user's posture and line of sight may also change. Therefore, it is required for the robot to coordinate its arm movements with the user's position, posture, and line of sight. However, the user's position changes dynamically. Hence, if the robot moves its arm in a dynamic fashion in response to the changing positions of the user, it becomes difficult for the user to view the display clearly (Figure 3). To overcome this issue, we propose a method in which the display can be adjusted by the user. This method necessitates that the robot adapt to the user in order to reduce the inconvenience caused to the user.

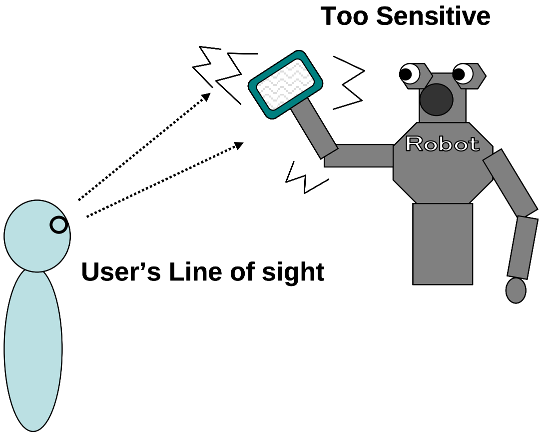

Figure 3 Abrupt Motions of the Robot

3.3.3 Playing Videos According to User's Requirements

The presenter robot needs to play tutorial videos that match the user's requirement. In order to implement this, the robot is first required to understand the user's requirements. In general, there are two natural approaches of exchanging knowledge, namely, speech and nonverbal communications such as joint attention. However, the presenter robot is a robot that it designed to assist the user; therefore, it is imperative that the system does not inconvenience the user as far as possible. Furthermore, there is a high possibility that the videos required by the user are related to the object nearest to the user (target object) at his/her work area. Therefore, it would be effective to select videos automatically in response to the positional relation between the user and target objects such as the parts of a bicycle. On the other hand, it is possible that different users performing the same task would desire to view different parts of the same tutorial video. Thus, it is essential for the robot to be able to distinguish between different user requirements.

3.4 TECHNICAL DEVELOPMENT PLAN FOR IMPLEMENTATION

3.4.1 Tutorial Video Production

In order that suitable tutorial videos can be selected automatically in response to the positional relation between the user and the targets, it is necessary to tag videos with meta-information such as position information in advance. Videos can be produced by a human or a listener robot. At present, it is difficult to directly use videos produced by the listener robot. These videos require to be converted by manual methods to be more easily viewable. Thus, the former method is appropriate for this study. Furthermore, it is important to prepare videos (using a zoom function) according to the specific task that the user intends to perform.

4 ROBUST AND STABLE ALGORITHMS FOR COORDINATING HUMAN-ROBOT INTERACTION

4.1 ARCHITECTURE OF THE PRESENTER ROBOT

The architecture of the presenter robot is shown in Figure 4. The presenter robot selects its behaviors in response to positional information acquired from motion capture sensors and touch information of the user acquired from touch sensors. In particular, the robot first calculates the positional relations, distance, and angle from the position information of the user and targets. Next, in response to the calculated positional relations, the robot decides upon a direction of movement and plays a video selected according to the nearest target. Additionally, the robot adjusts the position of its hand according to the touch information of the user. The robot pauses the video when the user touches its arm.

This architecture includes several hidden aspects of the implementation. These aspects concern the method of processing noises in motion capture data, detecting the user's nonverbal behaviors that strongly indicate his/her requirements from positional relation information, moving its wheel truck to a safe position, and adjusting the display position to adapt to the user.

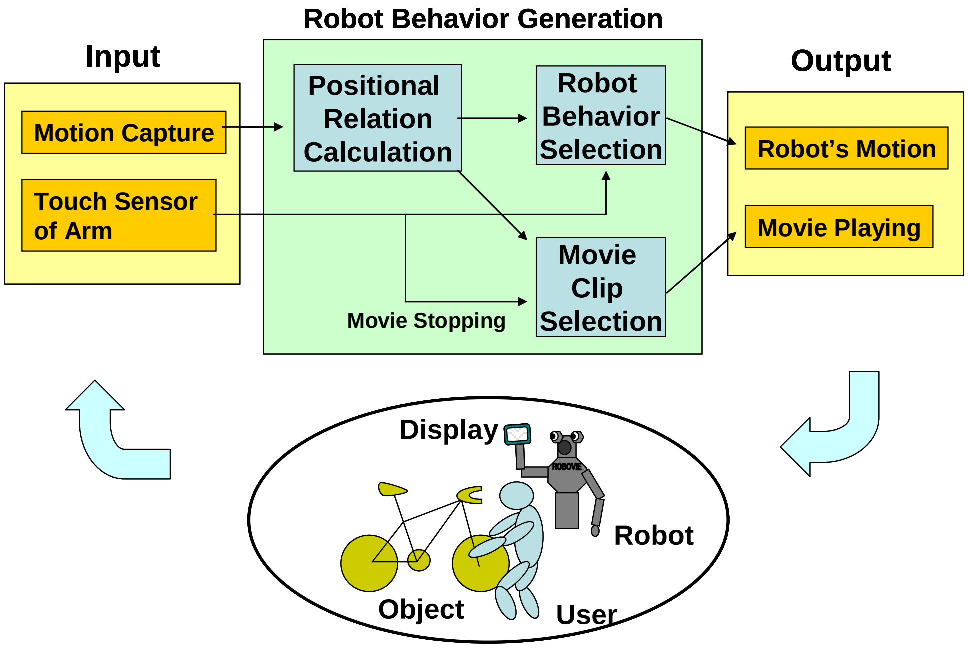

Figure 4 Architecture of the Presenter Robot

4.2 ROBUST AND STABLE BEHAVIOR GENERATION

In order to function in a noisy environment, the robot integrates its intentions and various information acquired from different sensors by using a Bayesian net. This method was employed by Ogasawara for the listener robot, and it can also be applied to the present study.

The networks used in our research are simple one-way graphs that are devoid of multiple routes, as shown in Figure5. The assumed event is calculated by applying the Belief Propagation algorithm at the time proportional to the number of nodes in the case where the evidence event is given. The details of the calculating expressions are as follows.

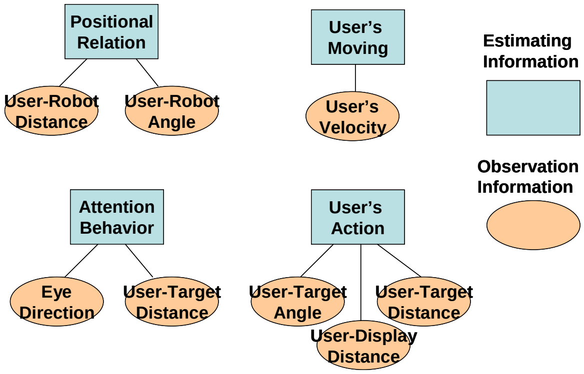

Figure 5 Bayesian Net Model

Node X is assumed, and node E is the evidence event. The part of E that coincides with X (precedent node) is called Ex_+, while that which is below X (led node by X) is called Ex_–. U, Y, and Z denote the nodes immediately below X, coinciding with X, and coinciding with Y (X is excluded), respectively. Moreover, the values of these nodes at the points where they can be obtained are assumed to be u, y, and z, respectively. At this point, confidence values (for Attention Behavior, User's Movement, User's Action, and Positional Relation, as shown in Figure 5) are calculated using expressions (1), (2), and (3) given below. “User’s Movement” refers to whether or not the user is moving, “User’s Action” refers to performing a task or adjusting the position of the display, while “Positional Relation” refers to the positional relation between the user and the robot.

$$\begin{aligned} & P(X \mid E) \\ & = P(X \mid E_{X\_-}, E_{X\_+}) \\ & = \frac{P(E_{X\_-} \mid X, E_{X\_+}) P(X \mid E_{X\_+})}{P(E_{X\_-} \mid X) P(X \mid E_{X\_+})} \\ & = \alpha P(E_{X\_-} \mid X) P(X \mid E_{X\_+}) \end{aligned} \tag{1}$$

(α is a regularized constant)

$$P(X \mid E_{X\_+}) = \sum_u P(X \mid u) \prod_i P(U_i \mid E_{Ui}) \tag{2}$$

$$\begin{aligned} & P(E_{X\_-} \mid X) \\ & = \beta \prod_i \sum_{Yi} P(E_{Yi\_-} \mid y_i) \sum_{Zi} P(y_i \mid X, z_i) \prod_j P(z_{ij} \mid E_{Zij}) \\ & = \beta \prod_i \sum_{Yi} P(E_{Yi\_-} \mid y_i) \end{aligned} \tag{3}$$

(β is a regularized constant)

For smooth interaction with the user, the robot may be required to react quickly, even before completing the calculation essential for making the correct decision. Therefore, we adopt a method involving the high speed repetition of a cyclic process that consists of a quick reaction and a correcting action to adapt to the user’s requirements. However, if this cyclic process is repeated very rapidly, the reactions of the robot become extremely sensitive, thereby making the robot unstable. In order to overcome this problem, it is necessary to attenuate high frequency signals. In other words, we plan to use a low-pass filter for processing the motion capture data.

In addition, a slow interaction that utilizes touch sensors and distance sensors will be combined with quick repetitive interactions; hence, a more stable action adjustment will be established. These sensitive and non-sensitive interactions will enable the robot to generate stable behaviors (smooth movements).

4.3 DETECTING TARGET OBJECT

In order to understand the user’s requirement (i.e., determining the video that the user wishes to see), the robot calculates the degree of attention given to each object when the user is performing a task and gazing at the individual objects, as shown in Figure 6. The confidence value for Attention Behavior can be calculated using expressions (1), (2), (3). Next, the attention value can be decided by adding weight to the confidence values of the objects. Finally, the object with the maximum attention value may be considered as the user’s target object.

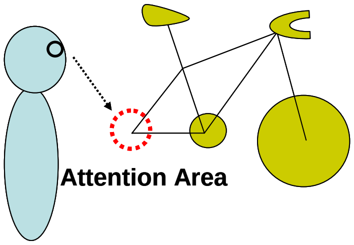

Figure 6 Focusing on an Object

The abovementioned approach can be explained using a detailed example. For the case in which the user is assembling the rear wheel, the robot calculates confidence values for User’s Movement, User’s Action, and Attention Behavior for each object. These values are 0.1 and 0.8, respectively, for the first two, while the confidence values of Attention Behavior for the objects, namely, the front wheel, rear wheel, front brake, and rear brake, are 0.3, 0.7, 0.2, and 0.5, respectively. By using these probabilistic values, the robot can perceive that the user is assembling the rear wheel.

4.4 STABLE TRUCK MOVEMENT AND SELECTION OF APPROPRIATE VIDEO

The robot is required to approach the user's work area and play the appropriate tutorial video, as shown in Figure 7. In order that the robot assumes the appropriate distance from the user, the distance D and direction F in relation to the user are calculated based on location information obtained using a motion capture technique. The center position of the robot is denoted as O (Ox, Oy, Oz), and the direction in which the robot is facing is denoted as A (Ax, Ay, Az). The center position of the user is denoted as U (Ux, Uy, Uz). The calculating expressions of D and F are as follows.

$$D = \sqrt{(O-U)^2} \tag{4}$$

$$F = \arccos\left(\frac{A \cdot (U-O)}{|A||U-O|}\right) \tag{5}$$

The robot moves via the wheel truck that it is placed on. The moving distance and direction of the truck are decided according to the expressions for D and F by using the Bayesian net and the threshold of the confidence value of Positional Relation of the user and robot.

The following example actively utilizes expressions (4) and (5). When D = 0.8 and F = –0.1 and the confidence values for the positional relations "It is away" and "Gap of the angle" are 0.7 and 0.2, respectively, the robot moves its truck and changes direction if these confidence values are above their respective thresholds.

In order to select suitable tutorial videos to match the user's requirements, the robot searches for the appropriate video from the previously tagged videos according to the object name that exhibits the maximum confidence value of User's Attention (4.3). In addition, the robot examines the duration of the attention time for the target object and accordingly selects other videos (with variable zoom functions) for the same target object. This method enables the robot to select videos automatically without requiring direct actions such as verbal instructions on the part of the user. Thus, this system can simplify the task being performed by the user.

Figure 7 Stable Display Position and Relevant Tutorial Video

4.5 ADJUSTING DISPLAY POSITION

Figure 8 shows a situation in which the user adjusts the position of the display in order to view it clearly. The presenter robot calculates the height of the user's head and waist by utilizing positional information obtained from sensors attached to the body of the user. We define H as the default height of the user from the average value of the head and waist heights and α and β (with an initial value of 0) as parameters for the user's adjustment value. The robot utilizes these parameters for deciding the position of the display for which the display height is H + α, display direction is β, and this position is assumed by the robot arm axis, role, and pitch, as shown in Figure 9.

A stable behavior is important, and the use of a Bayesian net proves to be effective in this regard. Specifically, the threshold of the User's Action confidence value is set in the network calculation; the display is kept stationary when the confidence value is above this threshold (this possibility is high when the user is working).

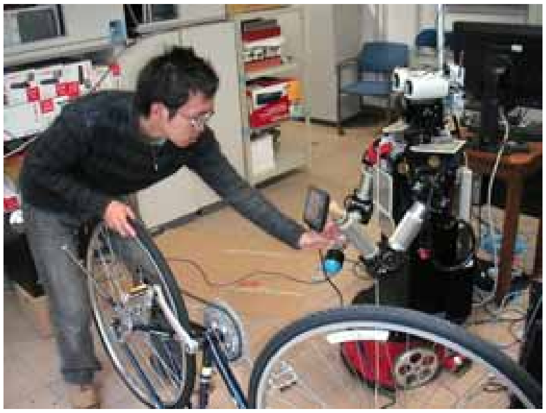

Figure 8 Adjusting Display Position

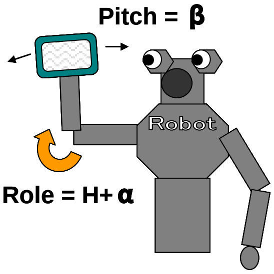

Figure 9 User Feedback Parameters

The ideal display position will change according to the user's height and posture, which varies among different individuals. Accordingly, the system also allows the user to directly adjust the position of the display by touching the touch sensors on the robot's arm. The robot can also change the position of the display by using the feedback parameters, α and β, of the user so that the user does not have to frequently perform such adjustments. This method enables the robot to gradually adapt to a user, thereby simplifying the task being performed by the user. Furthermore, the system can adapt to various users.

5 FUTURE PROSPECTS

5.1 PRESENT STATE OF IMPLEMENTATION

We have implemented modules such as approaching a target and adjusting the position of the display. Based on the architecture proposed in this paper, we plan to implement other modules for the presenter robot system. In the near future, we will discuss methods of evaluating the implemented system and conduct evaluation experiments.

5.2 POTENTIAL APPLICATION

In this paper, we discussed a presenter robot about a task of assembling/disassembling a bicycle as an example. But the ability to communicate knowledge to humans can be applied to various fields. For example, as an application of medical field the presenter robot can communicate knowledge, which are techniques of a doctor skilled in surgery, to a trainee doctor through tutorial video contents. The presenter robot can also be applied to other fields, sports, gardening, etc.

6 CONCLUSION

In this paper, we have demonstrated a presenter robot as a medium of knowledge that efficiently communicates information to humans. Further, we verified that the architecture of the presenter robot is stable and robust. In the near future, we will implement the presenter robot system on the basis of this architecture and evaluate it.

ACKNOWLEDGEMENT

We would like to thank reviewers whose comments about potential applications of the presenter robot are very useful.

REFERENCES

[1] Ito, M., Tani, J. (2004). Joint attention between a humanoid robot uses in imitation game. *Proceedings of 3rd International Conference on Development and learning*, 2004.

[2] Shiomi, M., Kanda, T., Eaton, D., Ishiguro, H. (2005). Communication Robot for Science Museum connected with Ubiquitous Sensor Network. *Interaction 2005*.

[3] Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N. (2004). Explorations in Engagement for Humans and Robots. *Artificial Intelligence May 2005*.

[4] Ogasawara, Y., Okamoto, M., Nakano, Y.I., Xu, Y., Nishida, T. (2005). How to Make Robot a Robust and Interactive Communicator. *KES 2005*, *LNAI 3683*, pages 289–295, 2005.