Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Development of User-Adaptive Value System of Learning Function using Interactive EC

Yuki SUGA * **Yoshinori IKUMA** * **Tetsuya OGATA** **
Shigeki SUGANO *

* *Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University, Tokyo, Japan. (Tel: +81-3-5286-3264; e-mail: ysuga, ikuma, sugano@sugano.mech.waseda.ac.jp)*
** *Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, Japan. (e-mail: ogata@i.kyoto-u.ac.jp)*

Abstract: Our goal is to create a user-adaptive communication-robot. We are developing a system for evaluating human-robot interactions. Although such evaluation is indispensable for learning algorithms, users' preferences are too difficult to model because they are subjective. In this study, we used the interactive evolutionary computation (IEC) to configure the value system of a learning communication-robot. The IEC is a genetic algorithm whose fitness function is performed by the user. In our experiment, we encoded the values of sensors (reward or punishment) into genes, and subjects interacted with the learning robot. Through the interaction, the subjects evaluated the robot by touching its sensors, and the robot learned appropriate combinations between input and output. Afterward, the subjects gave their scores to the experimenter, and the scores were regarded as the fitness values of the corresponding genes. These sequences were continued until the 4 generation, and then the subjects compared three of their best genes and two of the experimenter's. We found that the user-adaptive value system is suitable for the communication-robot.

1. INTRODUCTION

Our goal is to create a robot, which can communicate with people in various ways. We think that human-robot communication is essential not only for giving commands to the robot, but also for entertaining and relaxing its user, so this communication should have variety and flexibility. However, most robots can only communicate using the scenarios designed by their developers. Although scenario-based communication is practical as a type of context-sensitive communication, like speech, it lacks variety and flexibility.

On the contrary, we are interested in a behavior-based communication. From this aspect, various complex behaviors can be generated through mutual interaction between a robot and its surroundings. Although a behavior-based technique is not yet suitable for context-sensitive communication, such a technique would have much more variety and flexibility. However, even though entire scenarios are not described in the behavior-based communication, the robot's behaviors, which correspond to the specific sensory input, are defined by the designer *a priori*. Ideally, the robot's behaviors should be configured not by its designer but by the user to achieve user-friendly human-robot communication. Because most users do not have programming skills, the robot should be able to adapt to the users' preferences.

Machine-learning algorithms, such as a neural network, reinforcement learning, and genetic algorithms, are applied to achieve this behavior. A learning algorithm enables a robot to change its behaviors through interaction. However, we think that the evaluation is a significant problem for such a communication learning method. In such cases, psychological insight can be applied to address problems associated with subjectivity in human evaluation. Ishiguro et al. conducted a series of behavior adapatation experiments by using policy gradient reinforcement learning [13]. In their experiments, they hypothesized that a person's gaze, motion speed, and the distance between the person and the robot indicate how well the interaction is going. Their experiments were very successful because they designed the evaluation system carefully by considering psychological insight and the initial experimental results. However, human evaluation and interaction are subjective, so it is quite difficult to build a human evaluation model suitable to all users.

To address this problem, we used interactive evolutionary computatiaon (IEC) [5]. The IEC is an evolutionary computation, such as a genetic algorithm, but the fitness function is performed by the user, so we do not have to model the user's preferences, and we can use a machine-learning technique to solve the problems associated with the subjectivity. We have previously used IEC in a human-robot communication system [12]. In our previous experiments, the reactive controller of a communication-robot was configured with IEC by repeated interaction and evaluation. Although IEC can personalize a user's robot without intervention from the designer, this requires a lot of time, and the user must continue inputting fitness values as the robot continues learning.

We incorporated an individual learning function into the IEC-based communication system. The robot learns appropriate combinations of inputs from its user and outputs (its behaviors) after it is repeatedly given a reward (or a punishment) by its user, and IEC configures the value system, which assesses whether the behavior is successful by itself. In the next section, our proposed algorithm is described in detail. In section 3, the experimental settings are shown. In section 4, we show the experimental results, and the effectiveness of this algorithm is

discussed. Finally, we summarize this paper and describe our future work.

2. OUR PROPOSED ALGORITHM

In this study, two types of machine-learning techniques are used at the same time. One is the individual learning function, and the other is IEC. We describe each of the techniques in more detail, and then we show our proposed system.

2.1 Individually Learning Function

Figure 1 shows the overview of the learning function for our communication-robot. The system is constructed of two blocks, one is the motion-learning function, and the other is the value system.

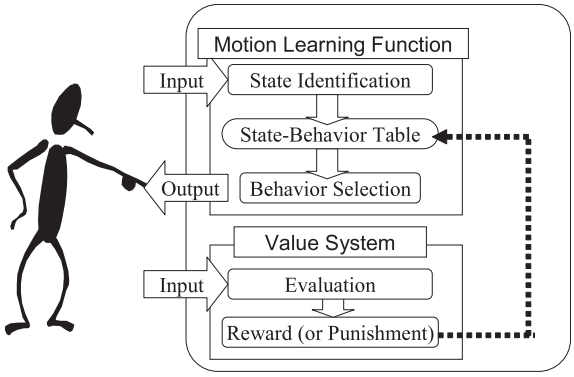


Fig. 1. Learning Function

If we regard a human-robot communication as an optimization problem, it would have a multi-peaked structure. Various behaviors can gratify the user, so the most appropriate motion cannot be defined. Therefore, the communication-learning function should be an unsupervised learning function, like reinforecement learning or a genetic alogrithm. A lot of researchers are developing such learning algorithms and some use them for human-robot interaction [11] [13] [15].

We are developing the value system, that is, a system for assessing whether the behaviors are successful. Many of these researchers have recently become interested in imitation learning [10] [14]. It resembles the way a mother teaches her child. Although this method of learning might be very plausible, we think that if a self-organizing learning system, which involves a self-supervised learning function, is installed into the robot, the robot can learn communicative motions through more natural communication.

To develop the value system for the communication-robot, we used the IEC that can deal with the problems associated with the subjectivity of people's preferences.

2.2 Interactive Evolutionary Computation

Interactive evolutionary computation (IEC) is an evolutionary computation (EC) whose fitness function is provided by human assessors. IEC, therefore, makes it possible to apply EC minimize the amount of subjectivity in people's preferences [5]. IEC has previously been applied to aesthetic areas, such as art and music. It has also been applied to robots [4].

We aim to use IEC for human-robot communication, because we believe that communication can only be learnt only by mutual interaction between a person and a robot, so our system consists of both of them. In our experiment, subjects interacted with a robot and simultaneously evaluated its response.

Our proposed algorithm is shown in Figure 2. First, the genetic pool is generated and the genes are initialized at random (1). Then, one gene is picked and translated into the parameters of the robot's value system (2). Next, the robot interacts with its user (3). Through the interaction, the value system generates the reward (or punishment) from the sensory inputs, and the motion-learning function estimates more suitable behaviors for communication (See Fig. 1). After the interaction, the user scores the robot, which is the fitness value of the corresponding gene (4). This sequence is continued until all genes are manually evaluated (5). Finally, genetic operators are applied to the genetic pool (6), and more appropriate genes are generated (7).

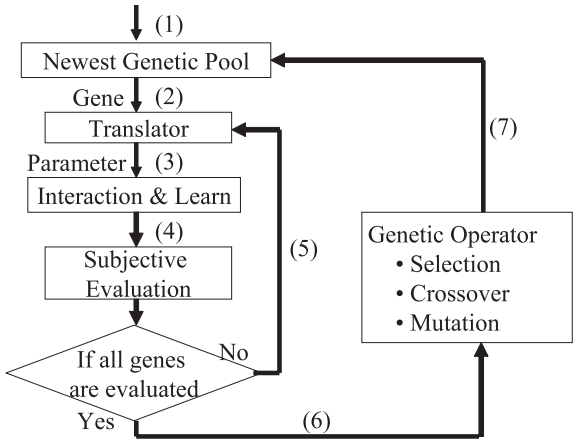


Fig. 2. Our Proposed Algorithm

3. EXPERIMENTAL SETTINGS

3.1 WAMOEBA-3

Our IEC experiment took a considerable amount of time because each subject evaluated a number of genes. Therefore, the robot must have a variety of behaviors to maintain the subject's interest. It must be harmless to people and be easy to maintain and customize. The robot must also be able to move without cables for the power supply or the control because cables prevent easy interaction with humans.

We used a communication-robot called Waseda Artificial Mind On Emotion BAse (WAMOEBA-3), as shown in Figure 3. WAMOEBA-3 is an independent, wheeled robot with built-in batteries and a PC. This robot was developed as a platform for communication experiments. Its upper body is analogous to a human one, and its size is about the average size of Japanese children: 825 mm wide, 1316 mm tall. WAMOEBA-3 weighs approximately 105 kg. It is equipped with two arms (7 degrees of freedom) and a head (8 degrees of freedom). Each joint has a torque sensor to measure the stress on the arm and the head. WAMOEBA-3 is also equipped with an omni-directional vehicle for locomotion, which can move in any direction without actually turning at any stage. This is advantageous for both the variety of its behaviors and for the safety of the subjects.

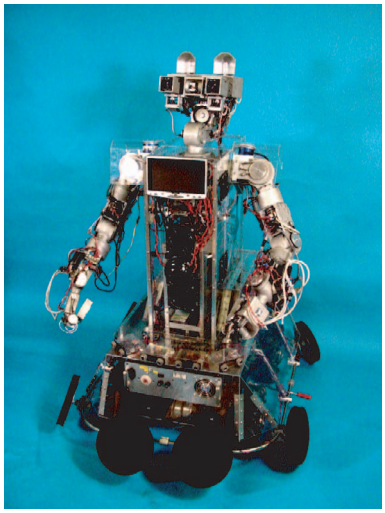


Fig. 3. WAMOEBA-3

WAMOEBA-3 also has a lot of sensors. It has shoulder covers into which 6-axis force sensors are installed to detect touches on the shoulders. Moreover, we installed pressure sensors into the top of its head and into both its hands. Its head has two CCD cameras and two microphones. Each camera can independently move vertically and horizontally, and each ear can rotate horizontally, which enable the robot to indicate the direction of its attention. The vehicle has eight bumper, three infrared, and eight ultrasonic sensors. Table 1 lists the specifications of the WAMOEBA-3 in more detail.

Table 1. Specifications of WAMOEBA-3

Dimensions	mm	1316 (H) x 825 (L) x 656 (W)
Total Weight	kg	105
Max speed	km/h	3.5
Payload	kgf/hand	5.0
Drive Time	hours	1.5
Drive Member	Camera DOF	1+1 x 2=3
	Ear DOF	2
	Neck DOF	3
	Vehicle DOF	3
	Arm DOF	6 x 2=12
	Hand DOF	1 x 2=2
Outside Sensors	Vision	CCD Color camera x 2 (x10 Optical zoom, x4 Digital zoom)
	Sound input	Microphone x 2 (Directional hearing, Voice recognition)
	Sound output	Speaker (Voice synthesis)
	Distance	Ultrasonic sensor x 8
	Human-like body	Pyroelectric sensor x 3
	Top of head	Pressure sensor x 3
	Joint stress	Torque sensor x 14
	Shoulder	6-Axis force sensor x 2
	Hand	Pressure sensor x 2
	Collision	Bumper switch x 8
Structural material		Extra super duralumin Titanium alloy (Ti-6Al-4V) Aluminum (52S)
CPU		Pentium 4 (3.2 GHz)
Microcomputers		VR5550-ATOM×5
OS		Linux
Battery		Lead-Acid Battery for EV

3.2 Behavior Learning Function

We implemented a very simple learning algorithm that is a reinforcement learning algorithm in which the policy is changed to maximize the reward given by the user after the interaction.

In this experiment, we prepared a simple table like Table 2. The table lists the values of behaviors corresponding to state conditions. First, the state condition, s, which is closest to the current condition of the robot, is selected from the table. Next, the behavior, b which has the largest value in the s row, is selected by priority. In this study, the best behavior is selected for a proportion, $1-\varepsilon$ (epsilon-greedy strategy). After that, the selected behavior is exhibited by the robot, and the user gives a reward (or a punishment) to the robot. Then, the behavior table is updated based on the reward. In this study, the value of the selected cell is updated by the folowing function:

$$Q_{t+1} = Q_t + \alpha(R - Q_t) \tag{1}$$

Here, Q is the value of the behavior, R is the reward (or the punishment), and α is the learning weight.

Table 2. Behavior Generation Table

	b_1	b_2	b_3	b_4
s_1	0.1	0.3	0.1	0.2
s_2	0.0	0.1	0.5	0.0
s_3	0.3	0.2	0.0	0.6
s_4	0.0	0.0	0.0	0.1

b_x: behavior
s_x: state condition

3.3 State Conditions, Behaviors, & Rewarding/Punishing Methods

In this experiment, we tested the adaptive value system of the learning function, so we had to determine whether the subjects could evaluate the teachability of the robot. On the basis of the preliminary experiments, we found that the subjects could not understand what should have been evaluated in a short amount of time, if the variety of the behaviors was too big. Therefore, we carefully defined six conditions and seven behaviors, as shown in Table 3. They are sufficiently clear for the subjects to imagine the appropreate combination of behaviors and state conditions.

Table 3. Behaviors and Conditions

Conditions	Sound from right
	Sound from left
	Camera detects something moving
	Touch on shoulder
	Touch on bumper
	No input, or other input
Behaviors	Ears turn right
	Ears turn to left
	Neck turns to right
	Neck turns to left
	Eyes look down
	Hands extend to subject
	Speaker makes sound “WAMOEBA”

We also carefully defined six methods for rewarding WAMOEBA as follows:

- Touch head
- Touch right shoulder

- Touch left shoulder
- Touch right hand
- Touch left hand
- Push bumper switches

In our value system, the reward (or the punishment) was given to the behavior learning function,when the sensors that correspond to the rewarding method described above were stimulated.

In this experiment, the subject could confirm whether the sensor he/she stimulated was a reward or a punishment by looking at the display mounted on the chest of the robot. If the sensor was a reward, the display was yellow, and if it was a punishment, the display was red.

These reward methods were too simple to satisfy the users, but they had sufficient variation to determine the diversity of the users' preferences, as we show in section 5.

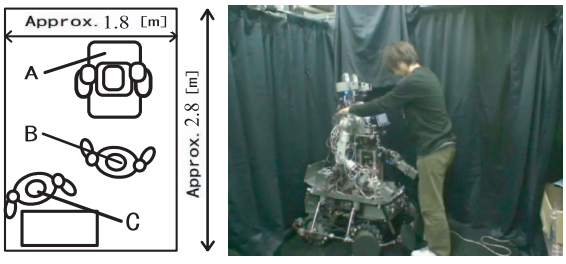
3.4 Genetic Settings

In this study, the reward/punishment values of the sensors are encoded into genes. Each gene has six elements, which correspond to the value of the sensors. Each element can be 1, 0, or -1 (A negative value means that the corresponding sensor is a punishment.). The probability of mutation is 10%, and if a mutation occurs, the element is increased or decreased randomly.

On the basis of previous IEC experiments [11] [12], we found that it is preferable that the experiment requires the subjects to concentrate for less than 3 hours at a time. Therefore, we used a population size of 10, and the IEC continued until the 4 generation.

3.5 Environment

The experiment was carried out in a small space in our laboratory (Figure 4). The robot (Figure 4a-A) and the subject (Figure 4a-B) interacted with each other until he/she wanted to stop. After that, the subject evaluated the teachability of the robot, and gave a score (the fitness value of the gene) to the experimenter (Figure 4a-C).



(a) (b)

Fig. 4. Experimental overview (a) and photograph of actual experiment (b)

3.6 Comparison of Created Robots

After the IEC experiment, the three best robots were selected and evaluated again to compare with two sample robots. We created the samples through the same IEC process.. Table 4 shows the genes of the samples. Touching them on the head and on the right shoulder were rewards, and touching them on the bumpers was a punishment.

Table 4. Value System of Our Prepared Robot

	Head	Shoulder Right	Shoulder Left	Hand Right	Hand Left	Bumper
Sample 1	0	1	0	0	0	-1
Sample 2	1	0	0	0	0	-1

They are very simple and intuitive value systems that we can easily imagine from the appearance of the robot, and we could observe them even during the subjects' IEC experiments.

4. RESULTS

In this study, we had two stages of the experiments. First, we used IEC to configure the value system of the individual learning function. Then, three of the three best optimized robots were compared with the two sample robots we prepared. Eleven subjects participated in our experiment. Nine of them are university students, and one of them is a woman. Five of them major in a scientific area, and three have programming experience.

4.1 IEC Experiment

The average experiment duration was 210 minutes.

Figure 5 shows the average fitness values of all subjects' experiments. The vertical axis indicates the fitness values and the horizontal axis indicates the generation. Although the fitness values increased slightly throughout the experiment, the diversity of the fitness values remained high.

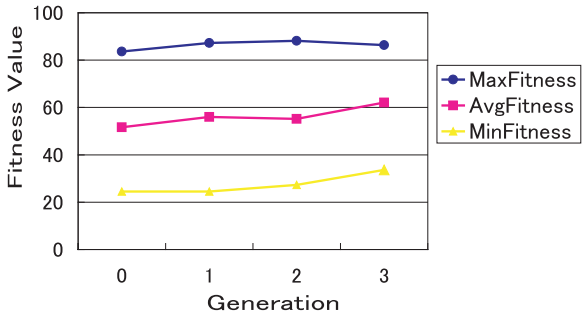


Fig. 5. Fitness Values

Table 5 shows the correlation between the fitness values and the values for genes of each subject in the last generation. A red cellmeans positive a correlation, and a blue cel means a negative correlation. A positive correlation indicates that the subjects gave high scores when touching the corresponding sensors was for a reward.

Subjects A, B, C, D, and E might have thought that touching the head should be a reward and touching the shoulder should be a punishment. Subject F prefered that toughing the shoulder should be a reward. On the contrary, subject K thought that touching the robot's head should be a punishment, and subjects G and H gave high scores to the robot whose right hand was touched as a reward.

Table 5. Correlation between Fitness (3 Generation) and Genes

Subject	Head	Shoulder Right	Shoulder Left	Hand Right	Hand Left	Bumper
A	0.92	-0.02	-0.61	-0.61	0.08	0.14
B	0.58	0.88	-0.92	-0.48	0.16	—
C	0.74	0.34	-0.34	0.00	0.27	-0.55
D	0.42	0.00	-0.44	-0.18	—	-0.84
E	0.36	0.05	-0.41	—	—	0.28
F	0.68	-0.10	0.26	0.06	—	-0.38
G	0.15	0.89	—	0.36	0.19	—
H	—	0.61	—	0.41	-0.35	0.02
I	0.03	0.06	0.29	0.49	-0.22	—
J	-0.12	0.08	—	-0.31	-0.39	-0.65
K	-0.49	0.78	0.48	0.03	0.40	-0.31

NOTE: A blank cell means that the correlation value cannot be calculated because all of the genes had the same values.

4.2 Comparison of Created Robots

Figure 6 shows the fitness values from the further experiments. IEC was performed on the best three optimized robots, and the sample robots contain the genes we prepared. We observed significant differences among IEC (1), and sample 1 ($P < 0.01$), and sample 2 ($P < 0.05$).

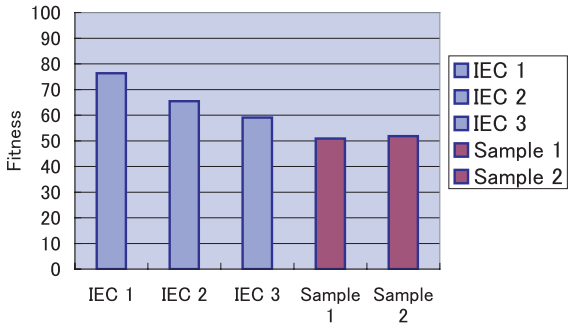


Fig. 6. Comparison of Experimental Fitness Values

5. DISCUSSION

For the IEC experiment, the amount of increase in the fitness values was insignificant. However, the amount of increase in the fitness value does not always mean that the learning is successful. Subjects evaluated the robots one-by-one, but they unconsciously compared the robots. Moreover, they would compare the robots during the span of one generation because we allowed the subjects to have a rest interval between generations. Therefore, mainatining a sufficient amount of diversity in the genetic pool is the most importantcriterion for evolutionary communication learning. If the amount of diversity is small, the genetic operator cannot work appropriately because estimating the successful gene pattern (scheme) becomes quite difficult.

In this study, the amount of diversity in the fitness values remained high throughout the experiment (See Figure 5), so we believe the subjects' evaluations of the robots were successful.

Next we analyzed the relation between the fitness values and the values of the genes. We observed slight tendencies, as we expected, in that touching the robot on the head was positively evaluated (reward), and touching the bumper switches of the robot was negatively evaluated (punishment). In spite of the small dimension of genes, however, at the same time we also observed variations in the results of the adaptation of the robot to each subject.

The expected results are easily obtained based on the robot's appearance. Because WAMOEBA-3 has a human-like upper body (See Fig. 3), most of the subjects gave high scores for the robot, which has the gene like that listed in Table 4 (Head = reward, bumper = punishment). If we had designed the robot's appearance more carefully, we could have obliged the users to interact with the robot in ways we expected. On the contrary, the shoulders and the hands had various values of genes. They were the difficult positions to design in the value system, and IEC is suitable for building value systems especially with these redundant sensors, which can reflect the user's own preferences.

We could confirm the effectiveness of the user-adaptive value system using IEC with a questionnaire survey which we carried out during the comparison experiments. The questionnaire items and the results of the factor analysis are shown in Table 6. As a result, three factors were obtained:

(1) Friendliness, Safety
(2) Understandableness, Interestingness
(3) Distinctiveness, Abiological

We also calculated the factor scores of the robot, which subjects trained by themselves (IEC), and ones we prepared (samples 1 and 2) (Table 7).

We observed a remarkable difference between IEC and the samples in factor 1 (Friendliness, Safety). Conversely, the difference in factor 2 (understandableness) was small. This means that, the IEC was not effective for gaining the understandableness of the robot's behaviors, but it made more optimized robots for the users, which engages the unique friendship between a robot and a person.

Table 6. Factor Loading

Adjective	Factor 1	Factor 2	Factor 3
Gentle	0.89	0.17	0.22
Familiar	0.84	0.42	0.17
Friendly	0.83	0.41	0.25
Communicating easily	0.83	0.47	0.12
Affable	0.82	0.29	0.25
Cordial	0.82	0.32	0.25
Warm	0.81	0.19	0.38
Amusing	0.79	0.51	0.18
Favorable	0.72	0.55	0.20
Intelligent	0.65	0.62	0.21
Safe	0.63	0.13	-0.01
Active	0.36	0.77	0.37
Understandable	0.47	0.67	0.29
Interesting	0.48	0.67	0.38
Funny	0.52	0.65	0.44
Abiological	0.12	0.40	0.88
Natural	0.23	0.31	0.86
Distinctive	0.25	0.52	0.69
Complicated	0.07	0.00	0.47

6. CONCLUSION AND FUTURE WORK

In this paper, we developed a user-adaptive communication system, which learns both individually and evolutionarily. We

Table 7. Factor Scores

	Factor 1 (Friendliness)	Factor 2 (Interestingness)	Factor 3 (Distinctiveness)
IEC	0.53	0.00	0.10
Sample 1	-0.37	0.11	-0.20
Sample 2	-0.18	-0.23	-0.29

used a very simple look-up table algorithm as an individual learning function. The behavior values are updated based on the reward/punishment generated by the value system. To develop the value system, we used interactive evolutionary computation (IEC). We carried out the experiment using a real robot named WAMOEBA-3, and we observed significant differences between the robots that subjects trained through IEC by themselves and the robots we developed. From an analysis of questionnaires, we found that IEC can improve the degree of friendliness of the robot as perceived by the users, by exploiting redundant degrees of freedom, which are difficult to design intuitively.

In our future work, we will use a more sophisticated motion generator. In this experiment, we defined the robot's behaviors *a priori*, but the behaviors also should be acquired through interaction with humans. Because the current look-up table algorithm cannot be applied to the continuous values, we think the neural network or actor-critic model will be promising.

ACKNOWLEDGEMENTS

This research was supported in part by a Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture, "the innovative research on symbiosis technologies for human and robots in the elderly dominated society", 21st Century Center of Excellence (COE) Program, Japan Society for the Promotion of Science.

REFERENCES

[1] Pfeifer, R., Scheier, C.: Understanding Intelligence; MIT Press (1999).

[2] Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Body Movement Analysis of Human-Robot Interaction; Proc. of Int'l Joint Conf. on Artificial Intelligence (IJCAI2003), pp. 177-182, 2003.

[3] Nolfi, S., Floreano, D.: Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines (Intelligent Robotics and Autonomous Agents); Bradford Books, 2000.

[4] Lund, H., Miglino, O., Pagliarini, L., Billard, A., Ijspeert, A.F Evolutionary Robotics - A Children's Game; IEEE Int'l Conf. on Evolutionary Computation (ICEC '98), pp.154-158 ,1998.

[5] Takagi, H.: Interactive Evolutionary Computation : Fusion of the Capabilities of EC Optimization and Human Evaluation; Proc. of the IEEE, Supecial Issue on Industrial Innovations Using Soft Computing, Vol. 89, No. 9, September, 2001.

[6] Holland, J.: Adaptation in Natural and Artificial System; MIT Press, 1992.

[7] Dawkins, R.: The Blind Watchmaker; Essex:Longman, 1986.

[8] Ogata, T., Sugano, S.: Emotional Communication Between Humans and the Autonomous Robot which has the Emotion Model; Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA'99), pp.3177-3182, 1999.

[9] Ogata, T., Matsunaga, M., Sugano, S., Tani, J.: Human Robot Collaboration Using Behavioral Primitives; Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS 2004), pp.1592-1597, Sept. 2004.

[10] Yokoya, R., Ogata, T., Tani, J., Komatani, K., and Okuno, G.H.: Experience Based Imitation Using RNNPB; Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006), pp.3669-3674, Beijing, China, Sep. 2006.

[11] Suga, Y., Ikuma, Y., Nagao, D., Ogata, T., Sugano, S.; Interactive Evolution of Human-Robot Communication in Real World; Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS2005), August, 2005

[12] Suga, Y., Endo, C., Kobayashi, D., Matsumoto, T., Ogata, T., Sugano, S.: User-Adaptive Human-Robot Interaction System using Interactive EC; Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006), pp.3663-3668, Beijing, China, Sep. 2006.

[13] Mitsunaga, M., Smith, C., Kanda, T., Ishiguro, H., Hagita, N.: Robot Behavior Adaptation for Human-Robot Interaction based on Policy Gradient Reinforcement Learning, in Proc. of the 2005 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, pp.1594-1601, 2005.

[14] Ogino, M., Toichi, H., Yoshikawa, Y., Asada, M.: Interaction rule learning with a human partner based on an imitation faculty with a simple visuo-motor mapping, Robotics and Autonomous Systems, Vol.54, No.5, pp.414–418, 2006.

[15] Inamura, T., Inaba, M., Inoue, H.: Incremental Acquisition of Behavior Decision Model based on Interaction between Human and Robots; Journal of the Robotics Society of Japan, Vol.19 No.8, 2001, (in Japanese).