Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Evaluating A Model for Generating Interactive Facial Expressions using Simple Recurrent Network

Yuki Matsui*, Masayoshi Kanoh†, Shohei Kato*, Tsuyoshi Nakamura* and Hidenori Itoh*
*Department of Computer Science and Engineering
Graduate School of Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
Email: {matui, shohey, tnaka, itoh}@juno.ics.nitech.ac.jp
†Department of Mechanics and Information Technology
School of Information Science and Technology, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota 470-0393, Japan
Email: mkanoh@sist.chukyo-u.ac.jp

***Abstract*—To improve face-to-face interaction with robots, we developed a model for generating interactive facial expressions by using a simple recurrent network (SRN). Conventional models for robot facial expression use predefined expressions, so only a limited number of expressions can be presented. This means that the expression may not match the interaction and that the person may find the expressions monotonous. Both problems can be overcome by generating expressions dynamically. We tested this model by incorporating it into a robot and comparing the expressions generated with those of a conventional model. The results demonstrated that using our model increases the diversity of face-to-face interaction with robots.**

***Index Terms*—Human-robot interaction, simple recurrent network, facial expression, emotion, Kansei robot, Ifbot**

I. Introduction

Several recently developed robots incorporate human-like features. For example, Repliee [1], [2] and SAYA [3], [4] resemble people. Kismet [5] generates facial expressions. Ifbot, the robot developed in our laboratory [6], incorporates sensibility communication technology that enables it to deal with facial expressions and emotions [7]- [12]. All of these robots generate facial expressions by using motors and are designed to communicate smoothly with people.

For robots to communicate smoothly with people, they need not only the ability to handle ordinary physical interactions but also to use *kansei* (*sensibility*). In other words, robots need to exhibit conciliatory behavior, and appear to enjoy the communication. They also need an interface to enable them follow the person's instructions. Moreover, robots should be able to interact with people on the basis of physical embodiment. As robots becoming more widely used, there will be a growing demand for them to act in a friendly manner and to use human-like communication methods. We have been working on various ways for robots to generate facial expressions as a component of their communication.

Facial expressions, as well as speech and gestures, play an important role in expressing emotions during human communication [13]. They can also play an important role in human-robot communication. They are especially important when for robots designed to entertain.

In general, robot facial expressions are generated by the intricate, coordinated movement of motors located in the robot's eyes, neck, and other areas. Since this requires much time and effort, the variety of facial expressions is limited. When a person expresses an emotion, he or she makes a facial expression with a certain pattern and features. However, the expression is always slightly different. If a robot's facial expression is always the same for a particular emotion and lacks variety, the person interacting with the robot will find it unnatural from our experience. For robots to express a variety of natural facial expressions, they need a way to dynamically generate expressions corresponding to emotions by synthesizing facial expressions using predefined expressions.

When people express an emotion, the emotion is based on a past state. A person has the flexibility to be stimulated by external agents, and emotions are generated by the transitions of the internal state. This leads to dynamic facial expression generation. To enable robots to exhibit dynamic facial expressions, we developed a model based on a simple recurrent network (SRN) [14] for generating facial expressions. An SRN generates output in accordance with previous state transitions. We implemented this model in the Ifbot robot and experimentally investigated its ability to improve human-robot communication.

II. Facial Expression Mechanism of Ifbot

Ifbot, shown in Figure 1, is 45 cm tall, weighs 9.5 kg, has two arms, and moves on wheels. The mechanism for controlling its facial expressions has 10 motors and 101 LEDs (Figure 2). The motors move Ifbot's neck, both sides of each eye, and both sides of each eyelid. The neck has two axes (θ_{N1}, θ_{N2}), and each side of each eyes has two axes (left: $\theta_{E1}^{(L)}$, $\theta_{E2}^{(L)}$; right: $\theta_{E1}^{(R)}$, $\theta_{E2}^{(R)}$). Each side of each eyelid has two axes (left: $\theta_{L1}^{(L)}$, $\theta_{L2}^{(L)}$; right: $\theta_{L1}^{(R)}$, $\theta_{L2}^{(R)}$). The LEDs work with the motors to express several emotions. They are located on the head (L_H), mouth (L_M), eyes (L_E), cheeks (L_C), and tear ducts (L_T). Those on the head can emit light in three colors

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

Fig. 1. Front and side views of Ifbot.

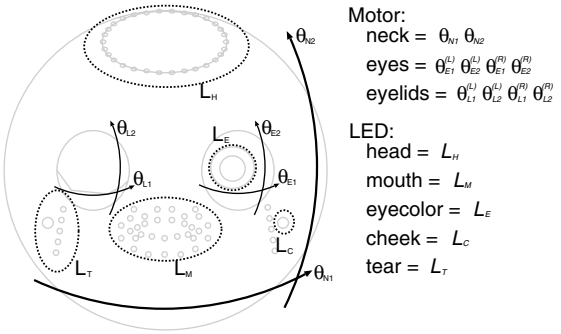


Fig. 2. Facial expression mechanisms of Ifbot.

(orange, green, red), those in the mouth can emit orange light, those in the eyes can emit light in three colors (green, red, blue), those in the cheeks can emit red light, and those in the tear ducts can emit blue light. Using this mechanism, Ifbot can generate various facial expressions.

III. Interactive Facial Expression Model

In general, a robot's facial expression is controlled by a motion file corresponding to the desired emotion. In other words, there is only one facial expression per emotion. Therefore, the facial expression for each emotion remains the same even over a series of interactions. Simply increasing the number of files will not eliminate the feeling of unnaturalness.

Our interactive model, which enables a robot to generate a unique facial expression in response to each stimulus, is based on a neural network. Neural networks are frequently used for mapping qualitatively different types of data, stimulations, and actuating parameters [16]. Neural networks can generalize the patterns learned during training to encompass new instances, and they are flexible. However, in general, a layered neural network cannot deal with time-series data. Therefore, we used a simple recurrent network [14], [15] as the basis of our expression-generation model because it can generate expressions on the basis of past state transitions.

A. Simple Recurrent Network

An SRN has a context layer between input layer and hidden layer. The number of units in the context layer equals that in

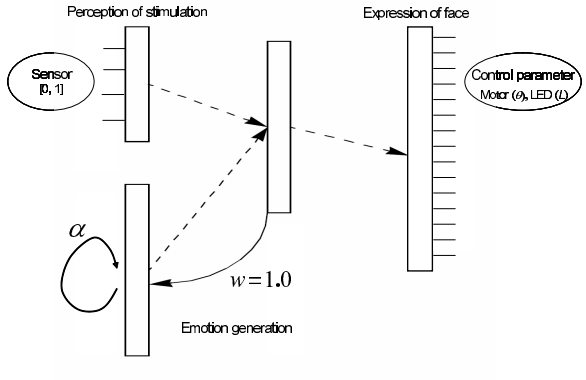


Fig. 3. Proposed interactive facial expression model.

the hidden layer because the units in both layers have a one-to-one correspondence. In an SRN, the activations are copied from the hidden layer to the context layer on a one-to-one basis, with a fixed weight of 1.0. The dotted lines in Figure 3 represent trainable connections. The data in the context layer at a given time reflects data from a previous hidden layer. Subsequently, the data in the context layer is returned to the hidden layer, where it is mixed with new input data. Thus, SRN deals with time-series data by adding a context layer.

B. Facial Expression Model using SRN

Our proposed model is illustrated in Figure 3. The input is a stimulation event, and the output is a facial expression. A context layer and hidden layer pair is an emotion for which the robot outputs a facial expression considering changes in past stimulations. The hidden layer contains an internal representation for mapping a facial expression to the stimulation event. The relationship between a stimulation event and the facial expression is mapped by learning the network. The robot can thus express a facial expression corresponding to the stimulation event.

This mapping of stimulation events to facial expressions enables an emotion to be connected with a facial expression. However, Cornelius states that emotion is a useful action to satisfy and free a sense and a desire. In addition, this action is thought to be acquired by the force of habit. For example, a cat will automatically assume a certain posture to protect its ears while fighting with another cat. This action came to be automatic in situations related to a threat [17]. The same kind of learning is needed for generating expressions in a robot that exhibit emotions. That is, repetitive learning of an action (facial expression) that is appropriate to the stimulation event is needed. The stimulation event and facial expression are simply mapped by learning the network. With this learning, various facial expressions can be produced by a robot.

This model of learned time-oriented relationships between basic stimulation events and facial expressions can generate various facial expressions using the generalization capability of

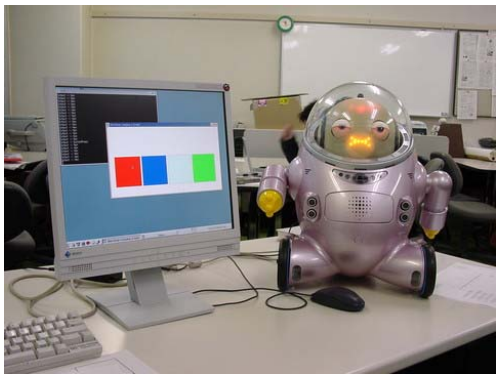


Fig. 4. Evaluation system using Ifbot.

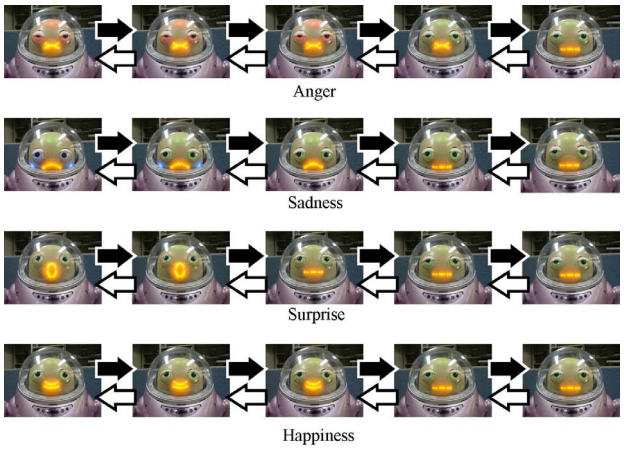


Fig. 5. Teaching data.

the neural network. When a stimulation event is input, the resulting facial expression is automatically generated because the network can learn and generate temporal patterns. Moreover, various mixed expressions can be synthesized corresponding to the timing of the stimulation event because the network has internal feedback. Without considering a specific interpolation method, the network can dynamically generate the facial expression.

We define the weight factor of the feedback connection with each context unit as α, and it remains fixed (Figure 3). The output of the context units at time t is given by

$$c_i(t) = \begin{cases} c_i(0) & (t=0) \\ \alpha c_i(t-1) + h_i(t-1) & (t>0) \end{cases}, \quad (1)$$

where $h_i(t)$ is the output of the hidden units at time t, and $h_i(0)$ and $c_i(0)$ are the initial values of the hidden units and context units, respectively. The learning of the network can be improved by setting an appropriate value of α.

IV. Evaluation System

A. Construction

The system used for the evaluation is illustrated in Figure 4. It is based on the assumption that a stimulation event can occur at any time. An emotion is identified for the event, and the robot, Ifbot, generates an appropriate facial expression. A person controls the interactions with the robot by using a display screen. As shown in Figure 4, red, blue, cyan, and green buttons are displayed. They represent four emotions (anger, sadness, surprise, and happiness) adopted from six basic emotions [18]. Red corresponds to anger, blue to sadness, cyan to surprise, and green to happiness. The system inputs a 1 to the unit in the input layer corresponding to the button pressed inputs a 0 when there is no stimulation event. The values of the facial expression control parameters in the model are used to control the robot. The 15 control parameters are expressed as

$$S = (\theta_{N1}, \theta_{N2}, \theta^{(L)}_{E1}, \theta^{(L)}_{E2}, \theta^{(R)}_{E1}, \theta^{(R)}_{E2}, \theta^{(L)}_{L1}, \theta^{(L)}_{L2}, \theta^{(R)}_{L1}, \theta^{(R)}_{L2}, L_H, L_M, L_E, L_C, L_T)^T, \quad (2)$$

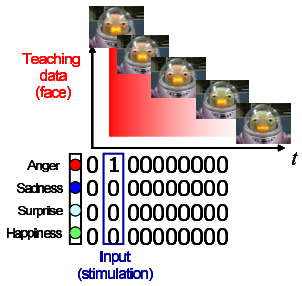


Fig. 6. *Proposed A.*

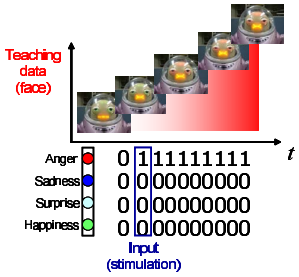


Fig. 7. *Proposed B.*

where $\theta^{(\cdot)}_{\cdot}$ are motor outputs corresponding to $\theta^{(\cdot)}_{\cdot}$, and $L_{\cdot}$ are patterns output from the LEDs of each part in Figure 1.

We used 4, 25, and 15 units in the input, hidden, and output layer, respectively (with 25 context units). The input units correspond to the four emotions. To train the network, we used four of Ifbot's facial expressions as teaching data. The teaching data was made of an emotional face and the default face by a linear interpolation. The temporal facial changes for these emotions are shown in Figure 5.

B. Characterization by Learning

We created four robots (two proposed and two conventional) that can generate expressions corresponding to the four emotions. The two proposed robots are referred to as "*Proposed A*" and "*Proposed B*", and the two conventional robots are referred to as "*Default A*" and "*Default B*".

We think that we can give the robot a character by the method of the training of stimulation and expression. We implement two characterized robots by learning in this paper.

1) **Proposed A:** *Proposed A* generated a facial expression for the appropriate emotion at the instant a stimulation event was input. The method used for training this robot (e.g. learning anger) is shown in Figure 6. For each input sequence, in which four bits were presented at a time, the correct output at the corresponding point in time is shown. At time $t = 1$, the input unit corresponding to the stimulation

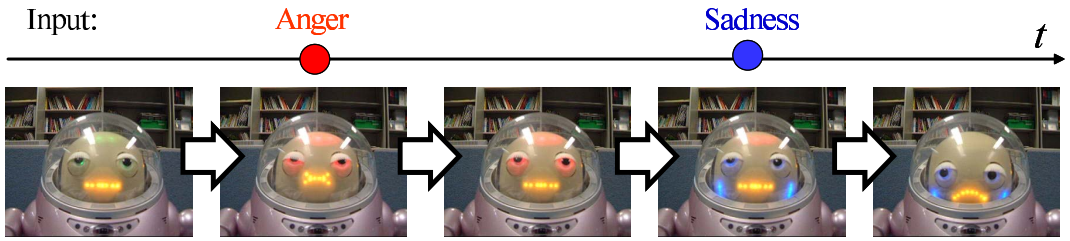


Fig. 8. Facial expressions using *Proposed A*.

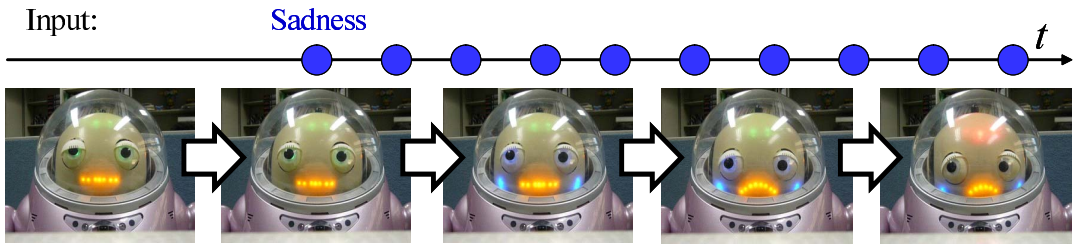


Fig. 9. Facial expressions using *Proposed B*.

event was input as 1.0, and the network trained the robot to generate the appropriate facial expression. For $t = 2 - 10$, the network gradually trained the robot to generate the default facial expression for when there was no stimulation event. Using this method, the network trained the robot to generate the appropriate facial expression for all four emotions using the right arrow facial change in Figure 5. In fact, when a stimulation event was input, the leftmost facial expression was generated.

2) **Proposed B:** *Proposed B* gradually generated a facial expression for the appropriate emotion as the stimulation event proceeded. The method used for training this robot (e.g. learning anger) is shown in Figure 7. At time $t = 1$, the input unit corresponding to the stimulation event as continuously input as 1.0, and the network gradually trained the robot to generate the appropriate facial expression. Using this method, the network trained the robot to generate all four emotions using the left arrow facial change in Figure 5.

Figure 8 and Figure 9 show examples of face changes using our proposed model. You can see the face of the fourth picture of Figure 8 that doesn't appear in the teaching facial data. Our proposed model can easily express the face with mixed emotions. Figure 9 shows that the robot generated emotional facial expressions by accumulating same stimulation.

Moreover, as follows, we prepared the *conventional robots* for facial expression. Conventional models for robot facial expression use predefined expressions.

3) **Default A:** *Default A* operated the same as *Proposed A*, except that it only generated facial expressions taken from the training data of right arrow face change.

4) **Default B:** *Default B* operated the same as *Proposed B*, except that it only generated facial expressions taken from the training data of left arrow face change.

V. Evaluation

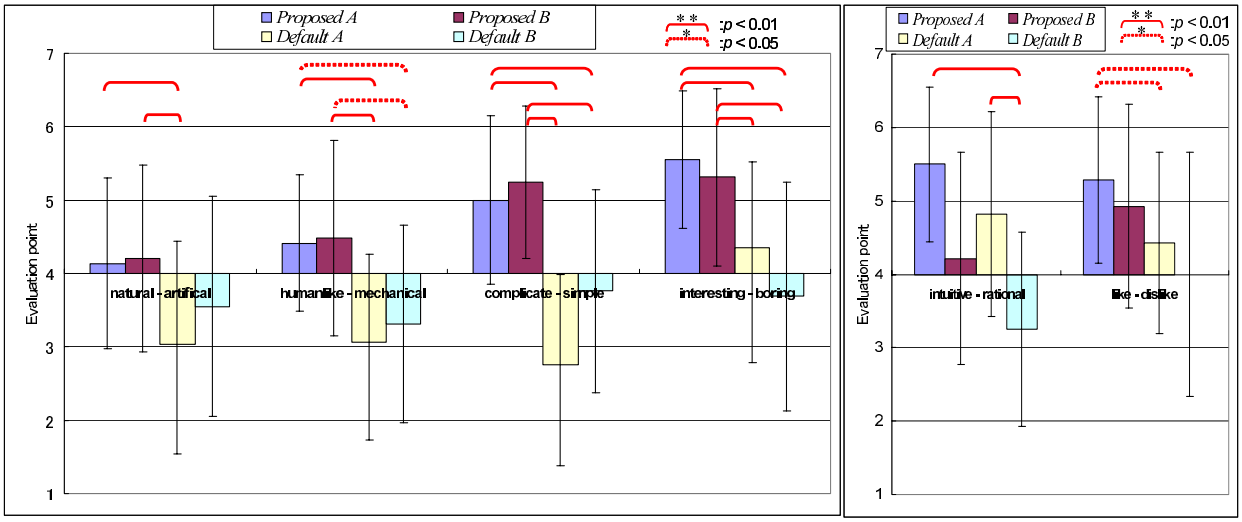
We subjectively evaluated the impression that each robot made on a person interacting with them in comparison with that made by the conventional system.

A. Procedure

The participants were 28 college students (20-24 years old, 23 men, 5 women). They interacted freely with the four robots (*Proposed A*, *Proposed B*, *Default A*, *Default B*) one robot at a time for as long as they wanted. In terms of content interaction, participants gave stimulus to the robot by using the button interface (e.g., input as shown in Figure 8 and Figure 9) and looked the facial expression of the robot. The interfaces for the robots were identical. The order in which the participants interacted with the robots was random for each person. At the end of the interaction with each robot, the participant completed a questionnaire. The evaluation was based on a semantic differential (SD) technique [19] in which values on a 7-point scale were assigned for six pairs of opposing evaluative adjectives:

- natural - artificial
- humanlike - mechanical
- complicated - simple
- interesting - boring
- intuitive - rational
- like - dislike.

For example, a score of 7 for the first pair meant that the robot acted completely naturally, while a score of 1 meant that it acted completely artificially. Therefore, the higher the score, the better the robot's performance. After evaluating all four robots, the participant ranked them in order of their



(a) Comparison of proposed system with default system (b) Evaluation for personal preference

Fig. 10. Experiment result of each system (robot) of each evaluation item.

overall performance and wrote a descriptive impression of the interactions.

B. Results

We tested the questionnaire using the Friedman test, a nonparametric one-way analysis of variance, and a Scheffe test, a test of statistical significance, because of ordinal scales. The mean values for the evaluated pairs are shown in Figure 10.

1) natural - artificial: The difference in evaluations between the two proposed robots and *Default A* was significant. Both proposed robots received a neutral evaluation, while the two conventional robots tended to receive an artificial evaluation. From the descriptive impressions, we found that *Default A* tended to react too quickly with little variety in its movements. This is attributed to the proposed method always changing the facial expression in accordance with the stimulation events and to the timing of the stimulation events. The proposed method can reduce the amount of artificial feeling in the interactions compared with the conventional method.

2) humanlike - mechanical: Both proposed robots were evaluated as humanlike, and both conventional ones were evaluated as mechanical; the difference was significant. We attribute this to the differences in the generated facial expressions between the proposed and conventional robots. The proposed robots displayed a variety of facial expressions because the expressions changed in accordance with the timing of the stimulation events. The proposed robots thus tended to create a less mechanical feeling. This was reflected in the descriptive impressions.

3) complicated - simple: The differences were significant for the all combinations of proposed and conventional robots ($p < 0.01$). Both proposed robots were evaluated as complicated, while both conventional ones tended to be evaluated as simple. Several opinions were expressed in the descriptive impressions: "The actions of the conventional robots felt monotonous"; "Since a mixture of expressions was expressed, the actions of the proposed robots felt complex". In short, the proposed robots were better able to generate expressions conveying mixed emotions.

4) interesting - boring: The differences were significant for the all combinations of proposed and conventional robots ($p < 0.01$). Both proposed robots were evaluated as interesting, while both conventional robots were evaluated as neutral. We consider that this evaluation item summarize the views on the above-mentioned evaluation items. These results suggest that the proposed model can increase the diversity of human-robot interaction.

5) intuitive - rational: This evaluation pair is significantly different from the others because it is not a comparison between the proposed and conventional robots but rather a characterization. *Robots A* (*Proposed A* and *Default A*) reacted instantly, while *Robots B* (*Proposed B* and *Default B*) generated facial expressions for each emotion by accumulating stimulation events. The participants generally thought that *Robots A* were intuitive and *Robots B* were rational. This result matches the setting of the characterization.

6) like - dislike: There was significant difference between *Proposed A* and conventional robots. However, as shown in Figure 10 (b), the tendency for a different preference depended on the person was seen for the control method and the characterization. From the descriptive impressions, some participants liked the quicker reactions of *Robots A*, and some liked the slower reactions *Robots B*. For example, one person thought that becoming angry quickly was more natural, while another thought that gradually becoming angry was more natural. Therefore, the difference for this pair is not significant because

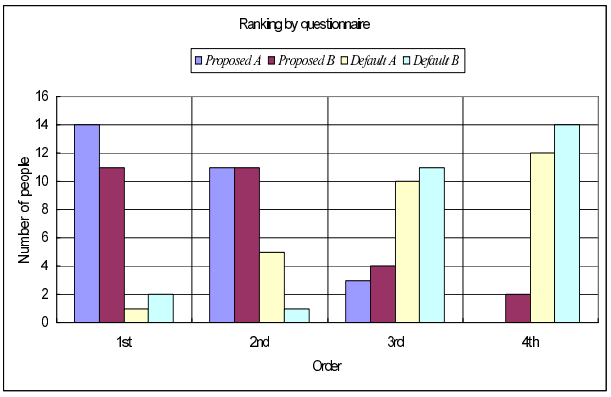


Fig. 11. Frequency distribution of each rank.

the evaluations varied a great deal.

C. Ranking

The results for ranking are shown in Figure 11 as a histogram. There was a clear difference: *Proposed A* and *B* were ranked first or second by virtually everyone. The reason for this is mostly explained by the results for like - dislike. Note that the other proposed method (characterization not suitable for the preference) was ranked second most frequently. This suggests that the good influences of the proposed method in the interaction are stronger than the benefits of the characterization.

VI. Conclusion

We have developed an interactive facial expression model that uses the feedback and generalization capabilities of a simple recurrent network. Only basic expressions are made and trained; the network uses them to automatically generate similar facial expressions. This reduces the time needed for generating predefined facial expressions. Moreover, the artificiality of and uncomfortable feeling generating by the robot's facial expressions are reduced. If a robot can add a feeling of emotion to its communication with a person, the person is more likely to develop a sense of intimacy with the robot. We focused on this advantage and examined it in a subjective evaluation. We implement the proposed model in a robot and evaluated the effectiveness of the dynamic facial expression during interaction between a person and the robot. The results suggest that dynamically generating facial expression using the proposed method gives the person interacting with the robot a better impression than that using a conventional method. We showed that using the proposed model

1) reduces the artificial and mechanical feeling created by the facial expressions of a robot,
2) a robot can express interest and complicated expressions by mixed emotions, and
3) a robot can be characterized by using facial expressions.

We used the frame of simplified interaction and evaluated it to judge the effectiveness of the proposed model because we wanted the participants to focus on the facial expressions of the robot during interaction. Considering symbiosis of the person, we need to evaluate the interactions in practical communications between a person and a robot.

Acknowledgment

Ifbot was developed as part of an industry-university joint research project among the Business Design Laboratory Co., Ltd., Brother Industries, Ltd., A.G.I. Inc., ROBOS Co., and the Nagoya Institute of Technology. We are grateful to all of them for their input.

This work was supported in part by Grant-in-Aid for Young Scientists (A) #20680014 of the Ministry of Education, Culture, Sports, Science and Technology, and Artificial Intelligence Research Promotion Foundation.

References

[1] T. Minato, M. Shimada, S. Itakura K. Lee and H. Ishiguro. Does Gaze Reveal the Human Likeness of an Android? *Proceedings of the 4th IEEE International Conference on Development and Learning*, pages 106–111, 2005.
[2] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro and N. Hagita. Android as a telecommunication medium with human like presence. *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI2007)*, 2007.
[3] H. Kobayashi, F. Hara, G. Uchida and M. Ohno. Study on Face Robot for Active Human Interface Mechanisms of Face Robot and Facial Expressions of 6 Basic Emotions. *JRSJ*, 12(1):155–163 (in Japanese).
[4] H. Kobayashi, F. Hara and A. Tange. A Basic Study on Dynamic Control of Facial Expressions for Face Robot. *Proceedings of IEEE International Workshop on Robot and Human Communication*, pages 168–173, 1994.
[5] C. Breazeal and B. Scassellati. A context-dependent attention system for a social robot. *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, pages 1146–1151, 1999.
[6] Business Design Laboratory Co. Ltd. *Communication Robot ifbot*. http://www.ifbot.net.
[7] S. Kato, S. Ohshiro, H. Itoh and K. Kimura. Development of a communication robot Ifbot. *The 2004 IEEE International Conference on Robotics and Automation (ICRA)*, pages 697–702, 2004.
[8] M. Kanoh, S. Kato and H. Itoh. Facial Expressions Using Emotional Space in Sensitivity Communication Robot "ifbot". *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1586–1591, 2004.
[9] M. Kanoh, S. Kato and H. Itoh. Analyzing Emotional Space in Sensitivity Communication Robot "ifbot". *The 8th Pacific Rim International Conference on Artificial Intelligence*, pages 991–992, 2004.
[10] M. Kanoh, S. Iwata, S. Kato and H. Itoh. Emotive Facial Expressions of Sensitivity Communication Robot "Ifbot". *Kansei Engineering International*, 5(3), pages 35–42, 2005.
[11] M. Gotoh, M. Kanoh, S. Kato, T. Kunitachi and H. Itoh", Face Generator for Sensibility Robot based on Emotional Regions. *The 36th International Symposium on Robotics*, 2005.
[12] H. Shibata, M. Kanoh, S. Kato and H. Itoh. A System for Converting Robot 'Emotion' into Facial Expressions. *IEEE International Conference on Robotics and Automation (ICRA 2006)*, pages 3660–3665, 2006.
[13] W. Von Raffler-Engel. Aspects of nonverbal communication. *Loyola Pr*, 1979.
[14] J.L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.
[15] J. L. Elman and E. A. Bates and Mark H. Johnson and Domenico Parisi and Kim Plunkett. Rethinking Innateness: A Connectionist Perspective on Development. *Bradford Books*, 1996.
[16] H. Yamada, K. Suzuki and S. Hashimoto. Interrelating physical feature of facial expression and its impression. *HCS2000-47*, 2001.
[17] Randolph R. Cornelius. The Science of Emotion; Research and Tradition in The Psychology of Emotion. *Prentice Hall College Div*, 1995.
[18] Paul Ekman. Unmasking the Face. *Prentice-Hall*, 1975.
[19] Snider, J. G., and Osgood, C. E. Semantic Differential Technique. *A Sourcebook*, 1969.