ORIGINAL ARTICLES

EMOTIVE FACIAL EXPRESSIONS OF SENSITIVITY COMMUNICATION ROBOT "IFBOT"

Masayoshi KANOH*, Susumu IWATA**, Shohei KATO** and Hidenori ITOH**

*School of Life System Science and Technology, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota 470-0393, Japan.
**Graduate School of Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan.

Abstruct: The "Ifbot" robot communicates with people by considering its own "emotions". Ifbot has many facial expressions to communicate enjoyment. We first attempted to extract characteristics of Ifbot's facial expressions by mapping these to its emotional space, which we discuss in this paper. We applied a five-layer perceptron to the extraction. We also propose a method of seamlessly changing facial expressions using the emotional space. We report on some results of facial changes obtained with the proposed method.
Keywords:*Sensitivity communication robot, Entertainment robot, Emotional space, Facial expressions*

1. INTRODUCTION

Recently, the robotics research field has been shifting from industrial to domestic applications, and several domestic robots have been developed [1–4]. As these robots require the ability to work in our daily lives, they also require interfaces to communicate actively with us. Communication involve emotions and thinking. That is, the psychological interaction between robots and people is very important for communication. Robots require several mechanisms to communicate actively:

- to recognize people's emotions.
- to have emotions themselves.
- to express their emotions through their face, and hand and body gestures.
- to describe their emotions in words.

To satisfy these requirements, we developed a sensitivity communication robot, the Ifbot [5–7], which communicates with people by understanding its interlocutor's voice, words and sentences, and expressing its emotions, moods and other feelings on its face. Ifbot received the 2003 Good Design Award #03A02002 in the product design section and in the amusement products and devices category from the Japan Industrial Design Promotion Organization.

Facial robots which can express its emotions have been developed, such as SAYA [8, 9], Repliee [10] and Kismet [11]. SAYA has humanlike skin, and creates facial expression faithfully. Repliee is a common name for two androids, a replica of a five-year-old Japanese girl (Repliee R1) and a woman android (Repliee Q1). Silicon skin covers whole each body of Repliee, and makes the skin feel humanlike. SAYA and Repliee are robots that closely resembles humans. On the other hand, Kismet has a simplified face that expresses emotions and interest. Ifbot also has a simplified face. Kismet makes its facial expressions with motors, but Ifbot makes them with motors and LEDs. Motors mounted on eyes, eyelids etc. and colors of LEDs create Ifbot's faces which express its internal emotions, purposes and so forth. Using various facial expressions, Ifbot can communicate with entertainment. In this paper, we describe the control of Ifbot's facial expressions considering its emotion.

All facial data incorporated into Ifbot start with a neutral face, go through expressive faces, and end with the neutral face. Therefore, Ifbot's face goes through the neutral face, when Ifbot express a facial expression created by connecting two facial sequences. On the other hand, human facial expressions in changing between two emotions, do not include neutral facial expressions. Ifbot requires smooth facial control for more naturally and entertaining expression. However, it is difficult to control Ifbot's facial-expression mechanisms directly, because it has many motors and LEDs. In this paper, we propose a method of seamless facial control that can direct Ifbot's fa-

Received April 26, 2004
Accepted June 2, 2005

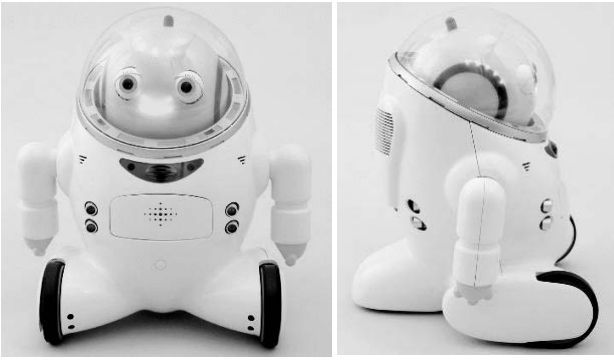

Figure 1: Front and side views of Ifbot.

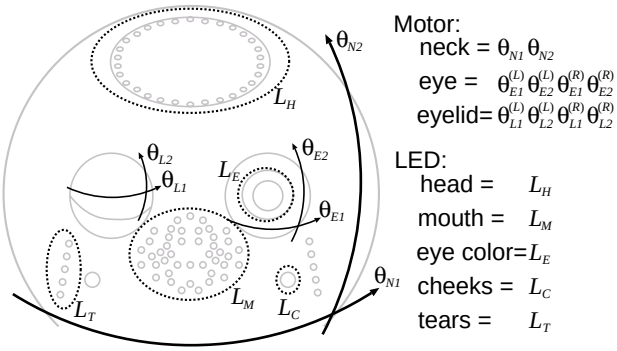

Figure 2: Facial-expression mechanisms for Ifbot.

cial expressions more easily by using its emotional model.

We consider creating *emotional space* for its emotional model. We first extract characteristics of Ifbot's facial expressions by mapping these expressions to its emotional space. We apply a five-layer perceptron to the extraction. The perceptron is an auto-associative neural network. Auto-associative neural networks effectively performs a non-linear principal component analysis. The emotional space with such a network represents difference between characteristics of two facial expressions as its distance. Using the emotional space, Ifbot can express more natural emotions on its face. Then, to create a natural impression for Ifbot's facial expressions, we propose method of seamlessly changing its facial expressions using the emotional space.

2. IFBOT

Figure 1 has a front and side view of Ifbot. It is 45 centimeter tall, weighs 7 kilograms, has two arms, moves on wheels. Ifbot communicates with its interlocutor using the emotions, which result in unique facial-expression mechanisms. Ifbot expresses its emotions, moods and other feelings on its face by using these mechanisms in communication.

3. FACIAL-EXPRESSION MECHANISMS OF IFBOT

Figure 2 outlines Ifbot's facial-expression mechanisms, which it has 10 motors and 104 LEDs. The motors actuate Ifbot's neck, and both sides of the eyes and eyelids. The neck has 2-axes (θ_{N1}, θ_{N2}), and each side of the eyes has 2-axes (left; $\theta_{E1}^{(L)}$, $\theta_{E2}^{(L)}$, right; $\theta_{E1}^{(R)}$, $\theta_{E2}^{(R)}$). Each side of the eyelids has 2-axes (left; $\theta_{L1}^{(L)}$, $\theta_{L2}^{(L)}$, right; $\theta_{L1}^{(R)}$, $\theta_{L2}^{(R)}$). The LEDs are set up for head, mouth, eye color, cheeks and tears. Using these mechanisms, Ifbot can communicate

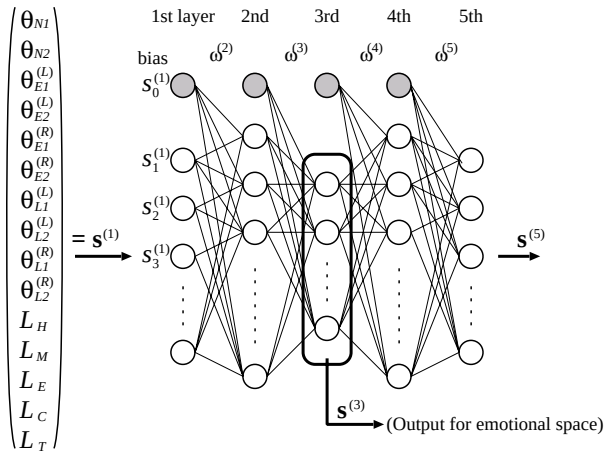

Figure 3: Creating emotional space.

with various facial expressions.

4. CREATING EMOTIONAL SPACE

Multi-layer neural networks can be used to do non-linear dimensionality reductions [12]. We apply a five-layer perceptron of the form in Figure 3 to extract the characteristics of Ifbot's facial expressions. The perceptron is an auto-associative neural network. The targets used to train the network are simply the input vectors themselves, so that the network attempted to map each input onto itself. Here, the input information is compressed by a process of dimensionality reduction, before it is regenerated to recover the original information. That is, the information is compressed, mixed and reorganized in the third layer. Similar faces, therefore, are classified as similar value in the third layer, when this network is used to extraction of facial characteristics. The authors of the article [13,14] have focused attention on this feature, and have analyzed and synthesized facial expressions of human beings. The authors used the third layer which extracts char-

acteristics of facial expressions as emotional space. We apply this idea to creating Ifbot's emotional space.

We use the following vector to input the perceptron:

$$\boldsymbol{s}^{(1)} = (\theta_{N1}, \theta_{N2}, \theta_{E1}^{(L)}, \theta_{E2}^{(L)}, \theta_{E1}^{(R)}, \theta_{E2}^{(R)}, \theta_{L1}^{(L)}, \theta_{L2}^{(L)}, \theta_{L1}^{(R)}, \theta_{L2}^{(R)}, L_H, L_M, L_E, L_C, L_T), \quad (1)$$

where, $\theta_{\cdot}^{(\cdot)}$ are motor outputs, and $L_{\cdot}$ are patterns outputted from the LEDs for head, mouth, eye color, cheeks and tears.

In the k-th layer, the j-th value $s_j^{(k)}$ of network output is given as:

$$s_j^{(k)} = f(u_j^{(k)}), \quad (2)$$

where $f(x)$ is a logistic sigmoid activation function, whose outputs lie in the range $(0, 1)$, and each $u_j^{(k)}$ is computed with the following equation:

$$u_j^{(k)} = \sum_i \omega_{ij}^{(k)} s_i^{(k-1)}. \quad (3)$$

Here, $\omega_{ij}^{(k)}$ is a weight, so that each unit computes the weighted sum of its inputs $s_i^{(k-1)}$. Note that we regard the bias parameters as being weights from extra input $u_0^k = 1$.

Sum-of-squares error E is given as:

$$E = \sum_i (s_i^{(1)} - s_i^{(5)})^2. \quad (4)$$

The perceptron network can be trained by minimizing error (back propagation):

$$\omega_{ij}^{(k)}(t+1) = \omega_{ij}^{(k)}(t) + \Delta\omega_{ij}^{(k)}(t), \quad (5)$$

where

$$\Delta\omega_{ij}^{(k)}(t) = \varepsilon d_j^{(k)} s_i^{(k-1)} + \eta \Delta\omega_{ij}^{(k)}(t-1) \text{ and} \quad (6)$$

$$d_j^{(k)} = \begin{cases} f'(u_j^{(k)}) \sum_l \omega_{jl}^{(k+1)}(t) d_l^{(k+1)} & (k \neq 5) \\ f'(u_j^{(k)})(s_j^{(1)} - s_j^{(k)}) & (k = 5). \end{cases} \quad (7)$$

ε is learning rate, and η is momentum in equation (6).

We used output $\boldsymbol{s}^{(3)}$ of the third layer of the perceptron to map Ifbot's emotional space.

5. EMOTIONAL SPACE IN IFBOT

We first prepared a questionnaire to analyze Ifbot's emotional space, in which we showed respondents 29 of Ifbot's facial sequences* , and they chose the best emotion corresponding each sequence. We provided seven options for classifying emotions: six basic emotions [15] (anger, disgust, fear, happiness, sadness, and surprise) and no classification. Figure 4 has examples of Ifbot's facial sequences. All facial sequences start with a neutral face, go through expressive faces, and end with the neutral face. We surveyed 50 people.

*All of the facial sequences used this questionnaire are incorporated into commercialized Ifbot. These are developed by designers. Each of sequences is created to project an impression.

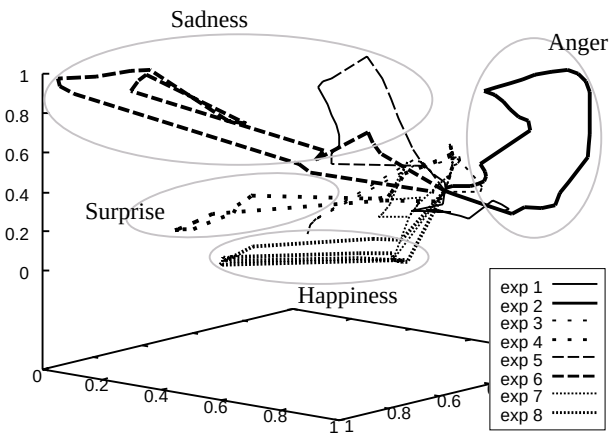

Figure 5: Emotional space for Ifbot. Emotional space is created from outputs of the third layer. The lines in this figure are the results of mapping Ifbot's facial sequences listed in Table 1.

Table 1 lists that each 2 facial sequences of anger, happiness, sadness, and surprise that were the most popular in the questionnaire. In the figure, the leftmost column denotes facial sequence number, and the values of the 2nd to 7th columns indicate support ratings when the sequence denoted in leftmost was shown. Note that disgust and fear did not earn a high support rating in the questionnaire.

We then created an emotional space in Ifbot using the five-layer perceptron described in the previous section. We used the values of 15, 45, 3, 45 and 15 for the number of units in each layer. Note that this type of network can extract characteristics of input data on the third layer, but cannot create meaning on its axes. We, therefore, chose the value of 3 for the number of third layer's units, to visualize and assess the emotional space. We evaluate the emotional space by confirming whether emotive facial sequences listed in Table 1 are classified on it. To train the perceptron network, we used the 29 facial sequences which were used in the questionnaire. Control values $\boldsymbol{s}^{(1)}$ in the sequences are used as inputs and targets of the network.

Figure 5 plots Ifbot's emotional space constructed by the network after the training of the network was completed. The lines in the figure are constructed by outputs

Facial sequence number: exp 4

Facial sequence number: exp 8

Figure 4: Ifbot's facial sequences.

Table 1: Results of questionnaire (%). This lists that each 2 facial sequences of anger, happiness, sadness, and surprise that were the most popular in the questionnaire. Note that disgust and fear did not earn a high support rating in the questionnaire.

Facial sequence no.	Recognized emotion (support ratings: %)						
	Anger	Disgust	Fear	Happiness	Sadness	Surprise	No class.
exp 1	**78**	14	0	0	4	0	4
exp 2	**84**	4	4	2	0	2	4
exp 3	0	0	0	14	0	**66**	20
exp 4	0	0	2	22	0	**72**	4
exp 5	2	6	4	0	**86**	0	2
exp 6	0	0	8	2	**90**	0	0
exp 7	0	4	0	**84**	0	6	6
exp 8	0	0	0	**96**	0	4	0

of the third layer of the network when Ifbot's facial sequences listed in Table 1 are inputted into the network. You can see that the facial expressions, which express Ifbot's emotions best, are classified in the emotional space.

6. CREATING FACIAL EXPRESSIONS

All facial sequences start with a neutral face, go through expressive faces, and end with the neutral face. Therefore, Ifbot's face goes through the neutral face, when Ifbot express a facial expression created by connecting two facial sequences. On the other hand, human facial expressions in changing between two emotions, do not include neutral facial expressions. Ifbot requires smooth facial control for more naturally and entertaining expression. However, it is difficult to control Ifbot's facial-expression mechanisms directly, because it has many motors and LEDs. In this paper, we propose a method of seamless facial control that can direct Ifbot's facial expressions more easily by using its emotional space.

We applied Gaussian convolution to creating seamless facial sequences, where two facial sequences were connected smoothly.

Although applying Fourier descriptors (e.g. [16–18]) is a good way of smoothing facial sequences, it has two main problems:

1. the number of constructing points for sequences is changed.
2. facial expressions may be lost when two facial sequences are smoothly connected, because the amount of smoothing is fixed, i.e. there is excessive smoothing.

In smoothing with Gaussian convolution, on the other hand, the number of the constructing points is not changed. Moreover, the amount of smoothing can be changed, because we can specify a different value for standard deviation σ in the Gaussian convolution for every constructing point. In this paper, by changing σ dynamically, we retain the activity of facial expressions and connect two facial sequences seamlessly.

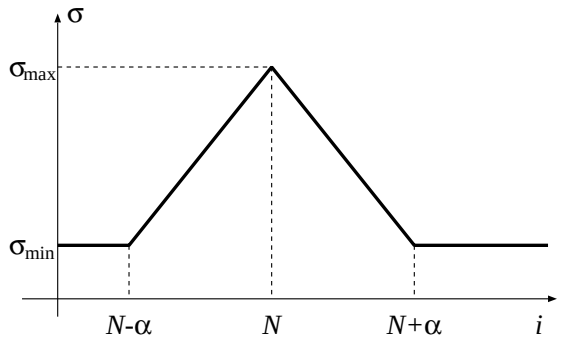

Figure 6: Dynamic change of σ.

First, let $\boldsymbol{s}_1(i)$ $(i = 0, \ldots, N-1)$ and $\boldsymbol{s}_2(j)$ $(j = 0, \ldots, M-1)$ be sequences in emotional space, then new sequence $\boldsymbol{s}_{1,2}(i)$ is given by connecting $\boldsymbol{s}_1(i)$ and $\boldsymbol{s}_2(i)$:

$$\boldsymbol{s}_{1,2}(i) = \begin{cases} \boldsymbol{s}_1(i) & (i < N) \\ \boldsymbol{s}_2(i-N) & (\text{otherwise}). \end{cases} \tag{8}$$

Second, the j-th element $s_{1,2}^{(j)}(i)$ of given sequence $\boldsymbol{s}_{1,2}(i)$ is convoluted in the following manner:

$$\begin{aligned} \mathcal{S}_{1,2}^{(j)}(i,\sigma) &= s_{1,2}^{(j)}(i) \otimes g(i,\sigma) \\ &= \int_{-\infty}^{+\infty} s_{1,2}^{(j)}(t) \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(i-t)^2}{2\sigma^2}} dt. \end{aligned} \tag{9}$$

If σ is fixed in the above convolution, it is a general smoothing method. We dynamically change the value of σ in the following manner:

$$\sigma = \begin{cases} -\dfrac{\sigma_{\max} - \sigma_{\min}}{\alpha}|i-N| + \sigma_{\max} & \\ & (|i-N| < \alpha) \\ \sigma_{\min} & (\text{otherwise}), \end{cases} \tag{10}$$

where $\sigma_{\max}$ is the maximum value of σ, $\sigma_{\min}$ is its minimum, and α is a smoothing parameter. How σ is dynamically changed is plotted in Figure 6. In the neighborhood of the neutral face, i.e. around N, a seamless facial expression is performed by increasing σ. To retain the activity of facial expressions, σ is decreased with moving from N.

Finally, Ifbot's facial sequence is reconstructed from $\boldsymbol{\mathcal{S}}_{1,2}(i,\sigma)$. The reconstruction is done by inputting $\boldsymbol{\mathcal{S}}_{1,2}(i,\sigma)$ to the third layer of the five-layer perceptron, and outputting Ifbot's facial sequence (see Figure 7). Note that the reconstruction does not use 1st and 2nd layers, in other words, does not use outputs from units in them.

7. VERIFYING EFFECTIVENESS OF PROPOSED METHOD

In this section, we discuss some experiments we did to verify the effectiveness of our proposed method. We used

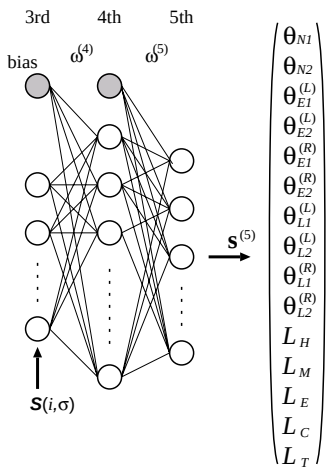

Figure 7: Reconstructing facial sequence from sequence in emotional space.

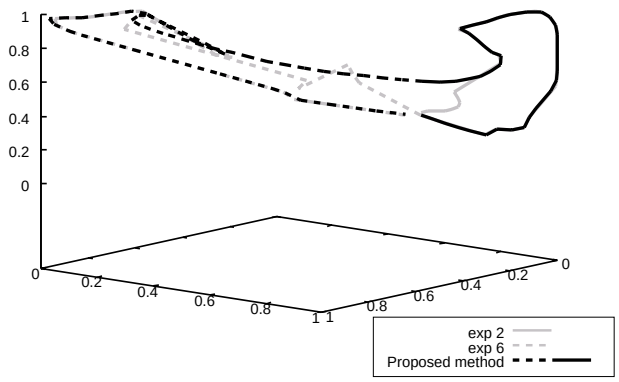

Figure 9: Sequence with proposed method.

$\sigma_{\max} = 50$, $\sigma_{\min} = 1$, and $\alpha = 150$ for the parameters in our method.

First, to verify whether facial expressions were seamlessly connected with our method, we compared it and an existing method that connects two facial sequences directly. Figure 8 shows facial change from anger (exp 2) to sadness (exp 6) with our proposed method. The facial expression is created by the sequence depicted by the black lines in Figure 9. Similarly, Figure 10 shows facial changes with the existing method, and the facial expression is created by the sequence depicted by the gray lines in Figure 9. In Figure 10, the neutral face (see (a)) is included between anger and sadness. On the other hand, in Figure 8, you can see that the face is not included and the seam is smooth. Here, in order to assess our method, we sent out a questionnaire: which facial change is more seamless, our method or the existing method? In the questionnaire, we used 10 facial changes with our method and

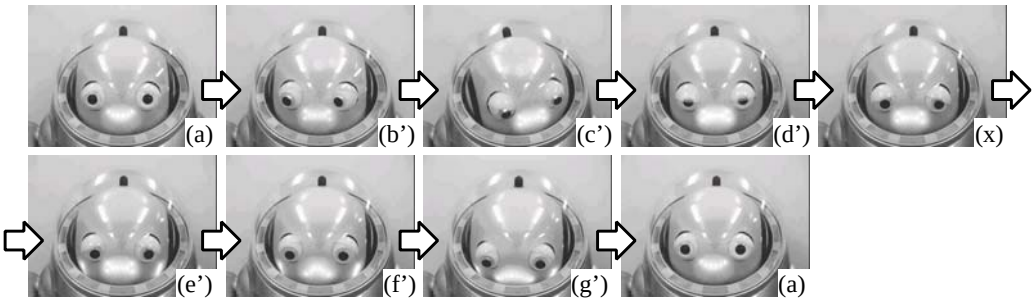

Figure 8: Facial changes from anger (exp 2) to sadness (exp 6) with proposed method.

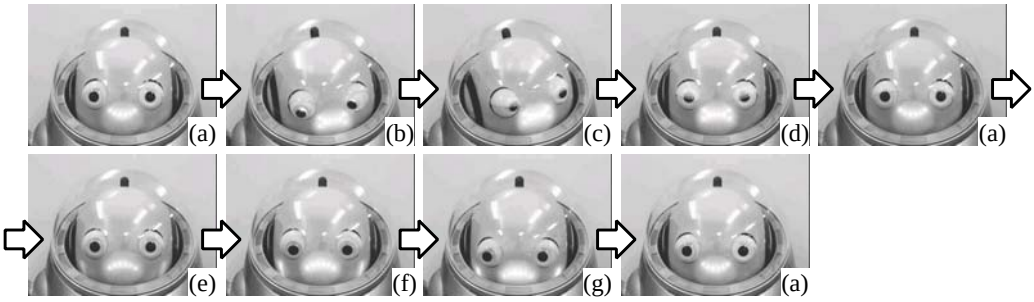

Figure 10: Facial changes from anger (exp 2) to sadness (exp 6) with existing method: (a) is neutral face, (b), (c) and (d) are facial expressions in exp 2, and (e), (f) and (g) are in exp6.

the existing method, respectively. We created the facial changes connecting two facial sequences which were chosen at random from 8 facial sequences listed in Table 1. We showed respondents two facial sequences with two method, and asked them which sequence is more seamless. We surveyed 20 people. As a result, our method was supported by an average 66% of them. In the questionnaire, the support raging of facial change in Figure 8 was 75 %.

Figure 11 and Figure 12 shows examples of facial changes with our proposed method. These support ratings were 70 and 80 %, respectively. You can see that these faces change seam and smooth.

To verify whether facial expressions had activity using our proposed method, we compared it and a σ-fixed method where the value of σ in the Gaussian convolution was fixed at 50. Figure 13 shows facial change from anger (exp 2) to sadness (exp 6) with the σ-fixed method. Here, to assess our method, we sent out a questionnaire: which facial change is more active, our method or the σ-fixed method? In the questionnaire, we used 10 facial changes with our method and the existing method, respectively. We created the facial changes connecting two facial sequences which were chosen at random from 8 facial sequences listed in Table 1. We showed respondents two facial sequences with two method, and asked them which sequence is more active. We surveyed 20 people. As a result, our method was supported by an average 72% of them. The result indicates that facial motion is smaller than with our proposed method, and activity is lost in facial expressions.

8. CONCLUSION

We attempted to extract the characteristics of Ifbot's facial expressions by mapping these expressions to its emotional space. We also proposed a method of seamlessly changing facial expressions using the emotional space. We reported on some interesting results for facial changes.

It is very important to consider whether we should design robot's facial and body mechanisms, and their motions. The essential challenging is a most important future work of us. We also do not focus attention on impression in around the facial expression connected with our method in this paper. We intend to survey what emotions are expressed by facial changes created by our method.

ACKNOWLEDGEMENTS

Ifbot has been produced by an industry-university joint research project, which was made up of the Business De-

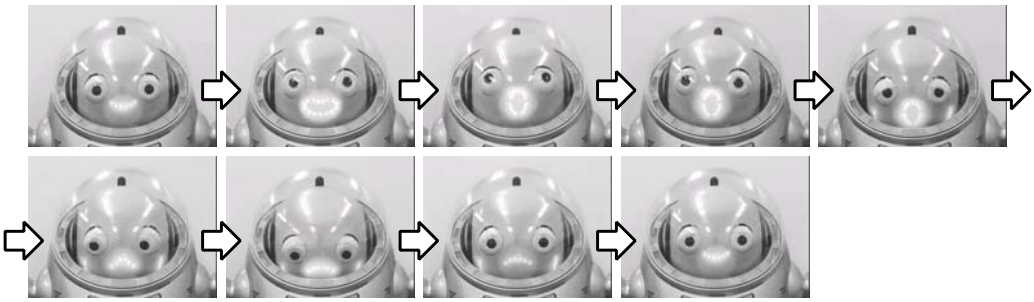

Figure 11: Facial changes from surprise (exp 4) to sadness (exp 6) with proposed method. Support rating of smoothness of this facial change is 70 %.

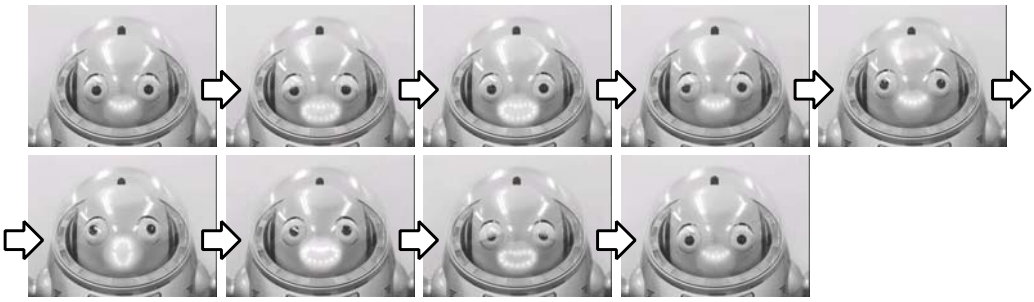

Figure 12: Facial changes from happiness (exp 8) to surprise (exp 4) with proposed method. Support rating of smoothness of this facial change is 80 %.

sign Laboratory Co., Ltd., Brother Industries, Ltd., ROBOS Co., and 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 (B) #17700160 of the Ministry of Education, Culture, Sports, Science and Technology, and the Kayamori Foundation of Informational Science Advancement.

REFERENCES

1. Y. Murase, Y. Yasukawa and K. Sakai et al.; Design of a Compact Humanoid Robot as a Platform, *The 19th Conf. of Robotics Society of Japan*; 789–790 (2001). (in Japanese).

2. Y. Kuroki, T. Ishida, K. Nagasaka and J. Yamaguchi; A Small Biped Walking Entertainment Robot SDR-4X with a Highly Integrated Motion Control, *The 20th Conf. of Robotics Society of Japan*; 1C34 (2002). (in Japanese).

3. M. Fujita, H. Kitano and T. Doi; Robot Entertainment, *In A.Druin and J.Hendler, eds.*; 37–70 (2000).

4. S. Kanda, Y. Murase and K. Fujioka; Internet-based Robot: Mobile Agent Robot of Next-generation (MARON-1), *Fujitsu*; 54 (4), 285–292 (2003). (in Japanese).

5. Business Design Laboratory Co. Ltd.; *Communication Robot Ifbot*. http://www.ifbot.net.

6. K. Tamura, S. Kato, T. Yamakita, K. Kimura and H. Itoh; A System of Dialogical Mental Health Care with Sensibility Technology Communication Robot, *International Symposium on Micromechatronics and Human Science*; 67–71 (2003).

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)*; 697–702 (2004).

8. 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*; 11 (1), 155–163 (1994). (in Japanese).

9. H. Kobayashi and F. Hara; Real Time Dynamic Control of 6 Basic Facial Expressions on Face Robot, *JRSJ*; 14 (5), 677–685 (1996). (in Japanese).

10. K. F. MacDorman and H. Ishiguro; The Study of

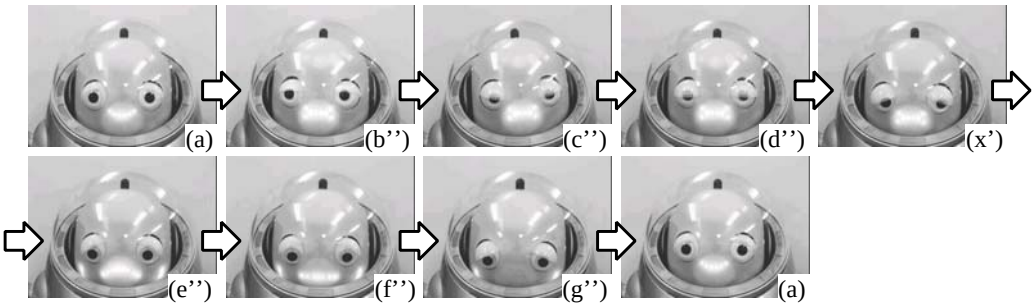

Figure 13: Facial changes from anger (exp 2) to sadness (exp 6) with σ-fixed method with σ fixed at 50.

Interaction Through the Development of Androids, *IPSJ SIG Technical Report CVIM*; 2004 (113), 69–75 (2004).

11. 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)*; 1146–1151 (1999).
12. C. M. Bishop; *Neural Networks for Pattern Recognition*, Oxford University Press (1995).
13. T. Sakaguchi, H. Yamada and S. Morishima; Construction and Evaluation of 3-D Emotion Space Based on Facial Image Analysis, *IEICE Transactions*; J80-A (8), 1279–1284 (1997). (in Japanese).
14. N. Ueki, S. Morishima, H. Yamada and H. Harashima; Expression Analysis/Synthesis System Based on Emotional Space Constructed by Multi-Layered Neural Network, *IEICE Transactions*; J77-DII (3), 573–582 (1994). (in Japanese).
15. P. Ekman; *Unmasking the Face*, Prentice-Hall (1975).
16. C. T. Zahn and R. Z. Roskies; Fourier Descriptors for plane closed curves, *IEEE Trans. Comp.*; C-21, 269–281 (1972).
17. G. H. Granlund; Fourier Preprocessing for Hand Print Character Recognition, *IEEE Trans. Comp.*; C-21, 195–201 (1972).
18. Y. Uesaka; Spectral analysis of form based on Fourier descriptors, *Science on Form, Proceedings of the First International Symposium for Science on Form*; 405–412 (1986).