Chapter 11
Emotion-oriented Human–Computer Interaction

Toshiyuki Yamashita[1], Ahmad Eibo, Takumi Ichimura, and Kazuya Mera

Abstract In the information-oriented society, people spend a lot of time interacting with computer systems. Therefore, we need to develop devices which make it a pleasure to interact with computer systems. In human face-to-face communication, nonverbal messages such as facial expressions, gestures, body movements, and postures can convey a speaker's emotions better than verbal messages. Among verbal messages, chiming in to express a listener's support and repeating the speaker's utterances can convey the listener's emotions, encourage the speaker to continue his/her talk, and make the conversation go smoothly. Therefore, we have developed several human–computer interactive systems using facial expressions as a human interface, and have constructed a human–computer interface which chimed in with the user and repeated the user's responses.

11.1 Introduction

Human face-to-face communication consists of both verbal messages and nonverbal messages. Nonverbal messages are composed of facial expressions, hand gestures, body movements, and postures, in addition to paralanguages such as tone, speaking rate, pause, and stumbling. Facial expressions in particular show an informative signature of a person's emotional state and persons can recognize emotions from another's facial expressions with great accuracy and consistency [1]. Therefore, attention to using facial expressions as human interfaces for human–computer interaction (HCI) has grown in recent years [2, 3, 4, 5]. In this chapter, we address our systems that can select and display the facial expression which expresses the user's emotions caused by HCI in order to make HCI pleasant.

[1] T. Yamashita
Tokyo Metropolitan University, 1-1-1, Minami-osawa, Hachioji, 192-0397 Tokyo, Japan
e-mail: yamashita-toshiyuki@center.tmu.ac.jp

Some kinds of verbal messages carry the same functions as nonverbal messages. Imagine a scene where two persons converse pleasantly. If one person suddenly stops replying to the other person, an unpleasant silence will fall upon them. In this human face-to-face communication, chiming in with remarks, repeating the other person's utterances, and so on, can sometimes encourage him/her to continue speaking, give pleasure to the communication, and smooth the conversation. This is also true of HCI. If the computer cannot understand a user's utterances and does not reply, the user will be discouraged and not want to continue the interaction. However, if the computer chimes in, or nods, or repeats the user's utterances, he/she will be encouraged to continue the interaction. For these reasons, we have constructed human interfaces which chime in with the user and repeat the user's utterances. In this chapter, we discuss the effects this has on HCI and describe our systems.

11.2 Facial Expressions as a Human Interface

In order to use faces for a pleasant HCI, we have developed both fuzzy reasoning [6, 7, 8] and neural network models [9, 10] for selecting the facial expression which expresses the emotions caused by several situations and have developed several applications using our models. First, we explain a face selection model by fuzzy reasoning and an application to e-mail.

11.2.1 Face Selection by Fuzzy Reasoning Model

Yamashita *et al.* [6, 7, 8] used nine facial expressions with three levels of brow deflection and three levels of mouth deflection, as shown in Figure 11.1.

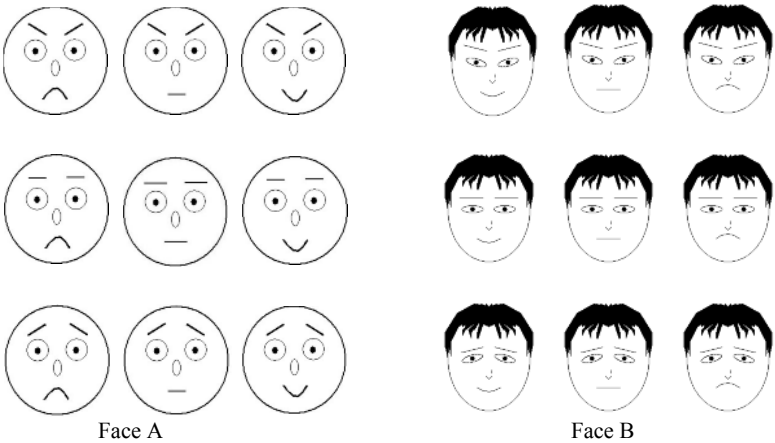


Figure 11.1 Nine facial expressions of Face A and Face B

Following Plutchik [11], we used the following eight emotions: “happiness”, “grief”, “anger”, “disgust”, “surprise”, “fear”, “anticipation”, and “resignation”.

We asked the subjects to select one or several facial expressions which expressed a given emotion. We assumed that the ratio of the subjects who selected the facial expression for a given emotion was the membership value in the fuzzy set of the facial expressions which expressed the emotion.

As shown in Figure 11.2, Yamashita *et al.* [6, 7, 8] proposed the following fuzzy reasoning model:

$$
\begin{array}{ll}
\text{Rule 1} & : A_1 \quad \Rightarrow P_1 \\
\text{Rule 2} & : A_2 \quad \Rightarrow P_2 \\
 & \quad \bullet \ \bullet \ \bullet \\
\text{Rule 8} & : A_8 \quad \Rightarrow P_8 \\
\text{Input} & : a_1 \text{ and } a_2 \text{ and } \ldots a_8 \\
\hline
\text{Conclusion} & : P',
\end{array}
$$

where $A_1, A_2, \ldots, A_8$ in the antecedent part are fuzzy sets of emotions, and P_1, P_2, …, P_8 in the consequent part are fuzzy sets of facial expressions which express a given emotion.

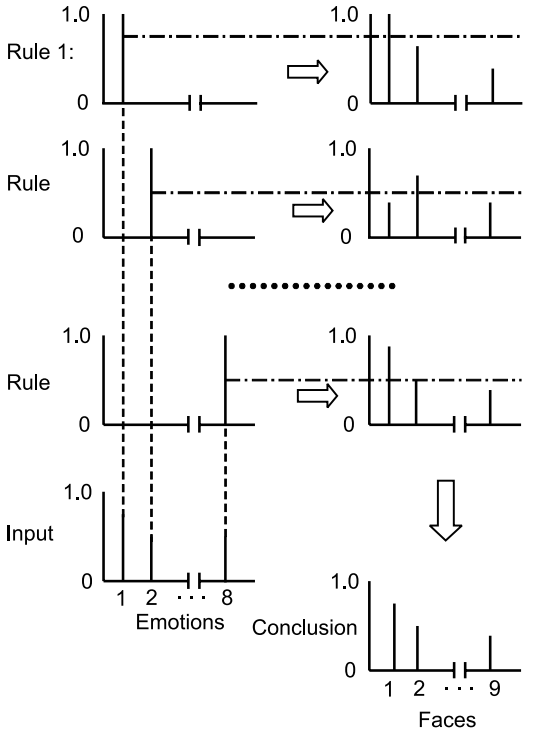


Figure 11.2 Fuzzy reasoning model for selecting a facial expression

If we have a_1, a_2, …, a_8 as the intensity of each emotion, then the fuzzy reasoning result of Rule i ($i = 1, 2, \ldots, 8$) is calculated by:

$$\mu_{Pi}, (z) = a_i \wedge \mu_{Pi}(z),$$

and the combined conclusion of Rule 1, Rule 2, …, Rule 8 is given as:

$$\mu_P, (z) = \mu_{P1}, (z) \vee \mu_{P2}, (z) \vee \ldots \vee \mu_{P8}, (z).$$

The facial expression with the highest or the second highest grade of membership should be selected as the facial expression which best expresses the emotions caused by a given situation.

11.2.2 An Application of the Model to e-mail

We applied our fuzzy reasoning model to an e-mail system, which could select and display the facial expression that expressed the emotions caused by the situation written in the sentences on the computer display.

Figure 11.3 shows an illustration of our system. First, the system receives sentences input by the user or from e-mail. Then it decomposes the sentences into words and transforms these words into original forms in terms of inflection [12]. Next, these words are compared to the words in the database for each emotion.

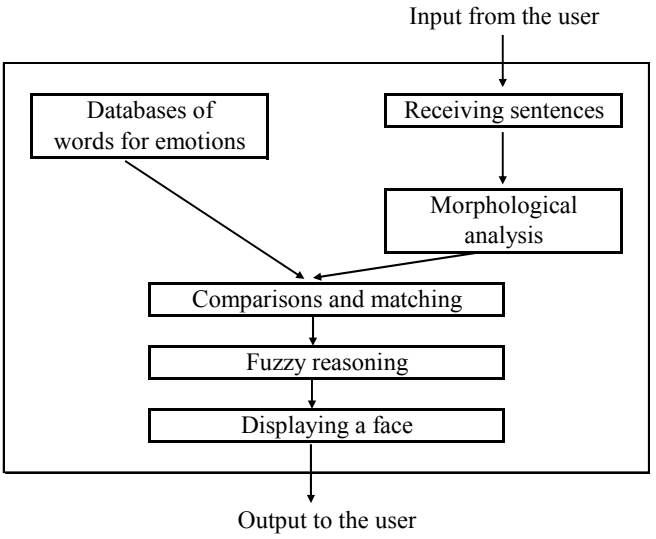


Figure 11.3 Overview of the system

For example, the sentences of e-mail in Figure 11.4 "I am very *glad* to hear that your son entered a junior high school this year. It seems only yesterday that I congratulated you on your son's entrance to an elementary school. I am *surprised* that children grow very quickly. I can visualize your *joyful* look. Congratulations!".

In these sentences, the first and third highlighted words "glad" and "joyful" are identified as representing the emotion "happiness". The second highlighted word "surprise" is identified as "surprise". Therefore, "surprise" has a membership value of 0.33 (= 1/3) and "happiness" has a membership value of 0.67 (= 2/3). These membership values are used as the input values for fuzzy reasoning. As a result of fuzzy reasoning, the facial expression with the highest membership value is selected and displayed, as shown in Figure 11.4.

Figure 11.5 shows an example of a computer display. Sentences are displayed on the upper side of the computer display, and a face is displayed below the sentences.

Our system can be applied to different interfaces. For example, Figure 11.6 shows our system used in two computer games [6, 8]. The left-hand side shows an example of the display for a computer game similar to the game Breakout. Since the player got a lower score than he expected, a facial expression showing his miserable feelings is displayed. The right-hand side shows an example of the display for a slot machine game. Since the player got the highest score, a facial expression showing happiness is displayed.

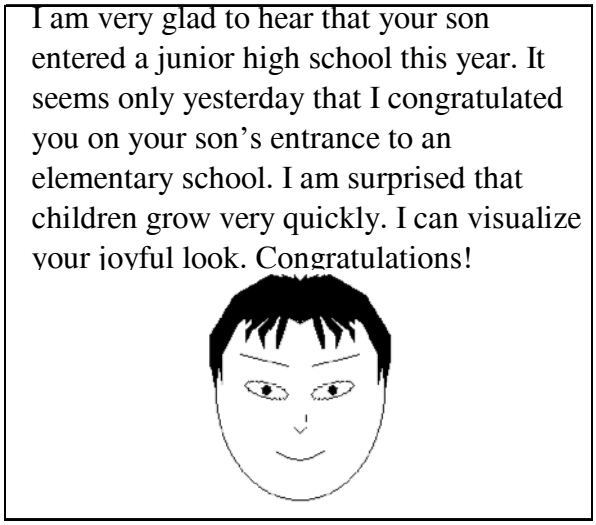


Figure 11.4 Example of e-mail display

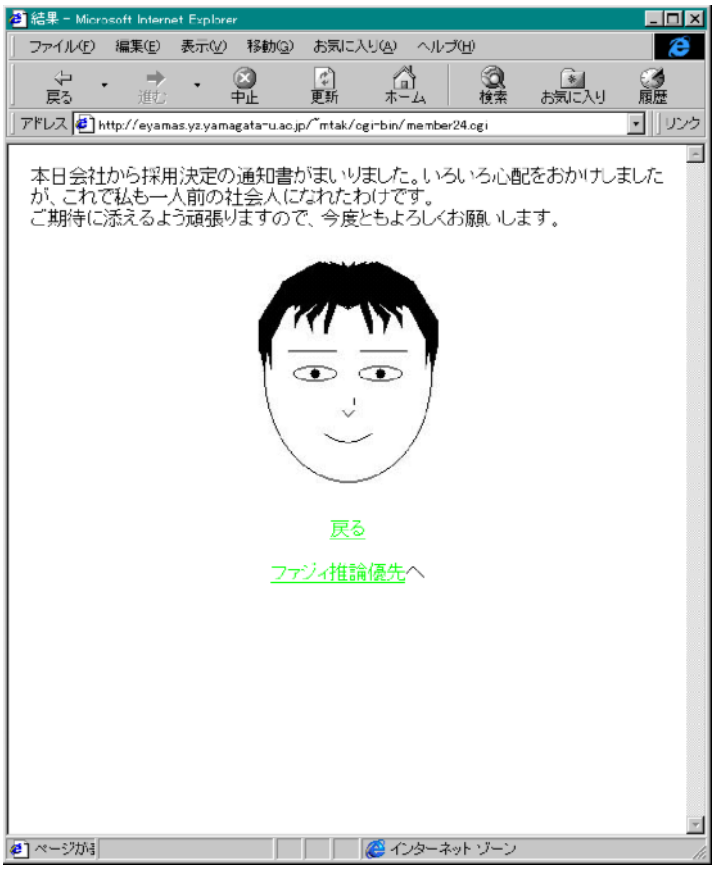
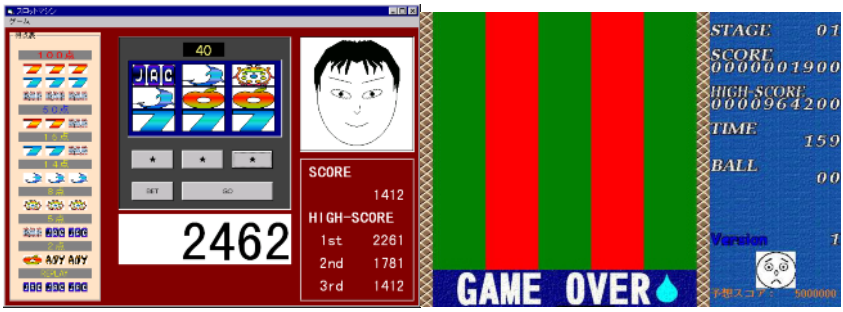


Figure 11.5 Example of a computer display



(a) (b)

Figure 11.6 Examples of applications to computer games: (a) the computer game Breakout, and (b) a slot machine game

11.2.3 Comparisons Among Faces

People often insert a so-called "face mark" ("smilies" or "emoticons") immediately after a particular sentence in an e-mail in order to convey the sender's emotions to the reader. As shown in Figure 11.7, the face marks are composed of simple characters. Although the face marks extremely simplify human facial expressions, they strongly convey the sender's emotions and make exchanging e-mails smoother.

Laughing: (^_^) (^o^) :) : -)

Be in a cold sweat : (^_^;) ^_^;; (^^;;

Be in tears: (; _ ;) (>_<) ; -< X -(

Dissatisfied: - (

Put out one's tongue: -P

Figure 11.7 Examples of face marks

We compared the effects of our faces with those of face marks in an e-mail [13]. We used three kinds of faces, that is, Face A and Face B as shown in Figure 11.1, and a face mark, for comparison. Moreover, we adopted two display conditions, that is, the one condition in which a relatively big face was displayed below the sentences, and another condition in which a small face was displayed right after a specific sentence. As shown Figure 11.8, three kinds of faces and two kinds of sizes were combined and produced six display conditions. We added the display condition without a face to the six display conditions. The subjects were asked to view each display and to rate the items such as "cannot obtain information quickly/can obtain information quickly", "unpleasant/pleasant", and "little information/much information" on a five-point scale ranging from 1 ("do not think so") to 5 ("think so"), with a score of 3 being neutral.

A factor analysis extracted three factors, that is, "ease of getting information", "pleasantness and familiarity", and "definition of information". The factor scores of the display conditions suggest the following: (1) facial expressions can add impression of "pleasantness and familiarity" to the information from the computer display. In particular, the small face mark, the big face mark, and small Face A can give the strong impression of "pleasantness and familiarity". (2) A small simple face mark added to a particular sentence can provide the user with the most information from an e-mail. (3) Showing the big face mark, big Face A, and big Face B below the sentences can give the impression that the sentences have "definite information".

I went to see a movie yesterday.
The movie was as amusing as I had expected. 
I recommend you to see this movie.

(a)

I went to see a movie yesterday. The movie
was as amusing as I had expected.
I recommend you to see this movie.

(b)

I went to see a movie yesterday. The movie was
as amusing as I had expected. (^_^)
I recommend you to see this movie.

(c)

I went to see a movie yesterday. The movie was
as amusing as I had expected.
I recommend you to see this movie.
(^_^)

(d)

Figure 11.8 Examples of e-mail sentences with faces: (a) an example of small Face A, (b) an example of big Face B, (c) an example of a small face mark, and (d) an example of a big face mark

11.2.4 User's Faces

We developed the system using user's faces. For generating the facial expression which expresses a given emotion, we used a sand glass type neural network as shown in Figure 11.9. A sand glass type neural network consists of two or more layered neural networks, which can learn some different teaching signals simultaneously and condense the input signals into hidden neurons and restore the original data from the hidden neurons [9, 10]. This section prepares the five-layer neural networks, which have 255 neurons in the first layer and fifth layer, 40 neurons in the second layer and fourth layer, and two neurons in the third layer. The third layers in each network are connected to each other and the neurons in the other layers activate independently in each network. The same teaching signals are set into the input and output neurons in each network to perform an identity mapping. The special feature of faces is compressed into the hidden neurons by Back Propa-

gation Learning. In particular, the two neurons in the third layer can depict the condensed features in the planner lattice after training the network.

The emotional facial expressions of volunteers are used as teaching signals. For each person there are six kinds of emotional facial expressions – "happiness", "sadness", "disgust", "anger", "fear", "surprise" – as shown in Figure 11.10. A facial expression model with emotion is constructed by the neural networks so that an emotion space appears in the third layer of the neural network. Figure 11.11 shows an example of the application of our model to e-mail.

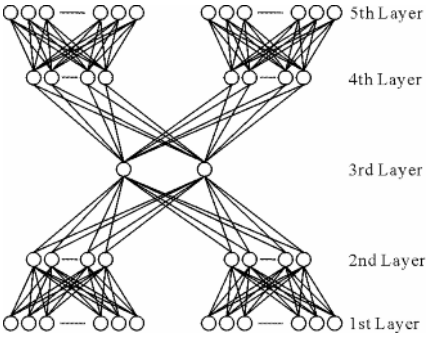


Figure 11.9 An overview of a sand glass neural network

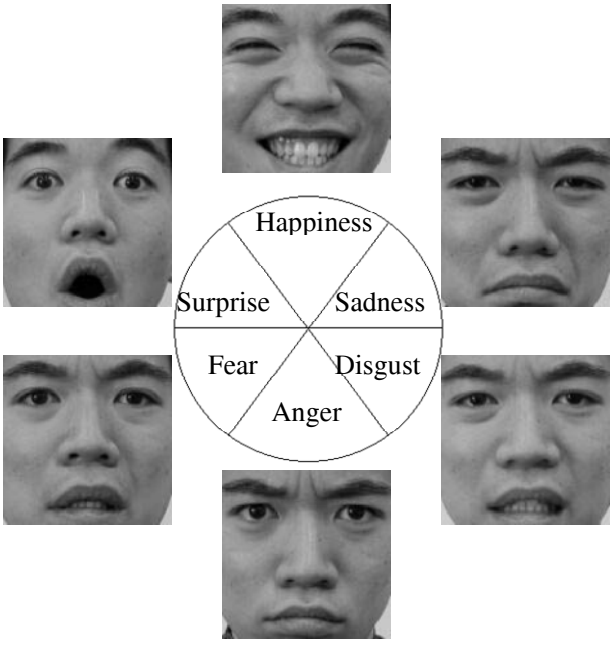


Figure 11.10 Emotion circle of a user's face

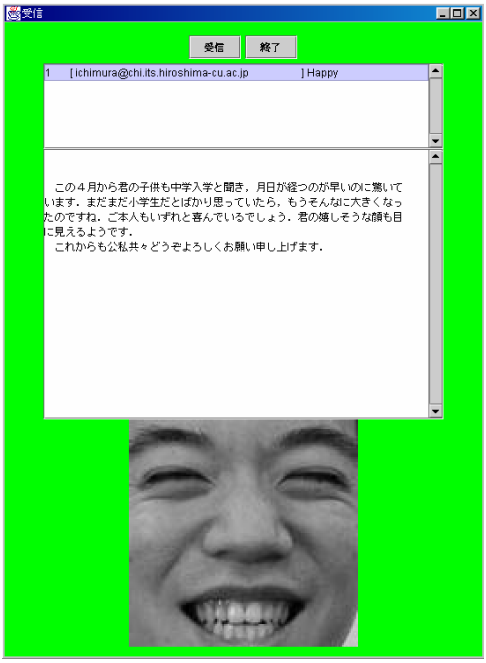


Figure 11.11 Example of an e-mail with a user's face

11.2.5 Noh Mask

Noh masks used in Noh play, which is one of the most popular traditional arts in Japan, are artificial and sometimes ambiguous, and can be interpreted as expressing various emotions [14, 15, 16]. That is, as the Noh mask turns upward and downward or to the left and right when the actor with the Noh mask moves on the stage, the audience can read diverse emotions from the Noh mask from the changing of the angle and the combination of light and shadow. It is usually said that an upward turn of the Noh mask, which is called *terasu* (shining), represents a pleasant and cheerful state of mind, and a downward turn of the Noh mask, which is called *kumorasu* (clouding), represents a gloomy state of mind. The audience enjoys reading delicate emotions from the Noh mask.

People can read various scenarios for Noh play from the Internet without going to Noh theaters, but they cannot enjoy reading the delicate emotions from the Noh mask. We applied the fuzzy reasoning model to selecting the image of the Noh mask which best expresses the emotions caused by some scenario in the Noh play [10]. Moreover, we constructed a computer display system which can allow users to read the scenario of the Noh play and enjoy the delicate emotions that the Noh mask expresses. We used *Koomote* as the Noh mask, because *Koomote* is one of

the most popular female masks and is known to have the richest variety of facial expressions. Figure 11.12 shows typical *Koomote* images corresponding to Ekman and Friesen's six basic emotions of "happiness", "sadness", "anger", "disgust", "surprise", and "fear" [17].

We illustrate our system using *Hagoromo* as a Noh play. The story of *Hagoromo* is as follows. A fisherman finds an angel's cloak on a branch of a pine tree on the beach. He steals it and so prevents her from returning to heaven. The angel is very sad because she cannot go back to the sky. After her entreaties, the fisherman decides to give the cloak back to the angel. The angel is very glad and dances to show the glory of heaven.

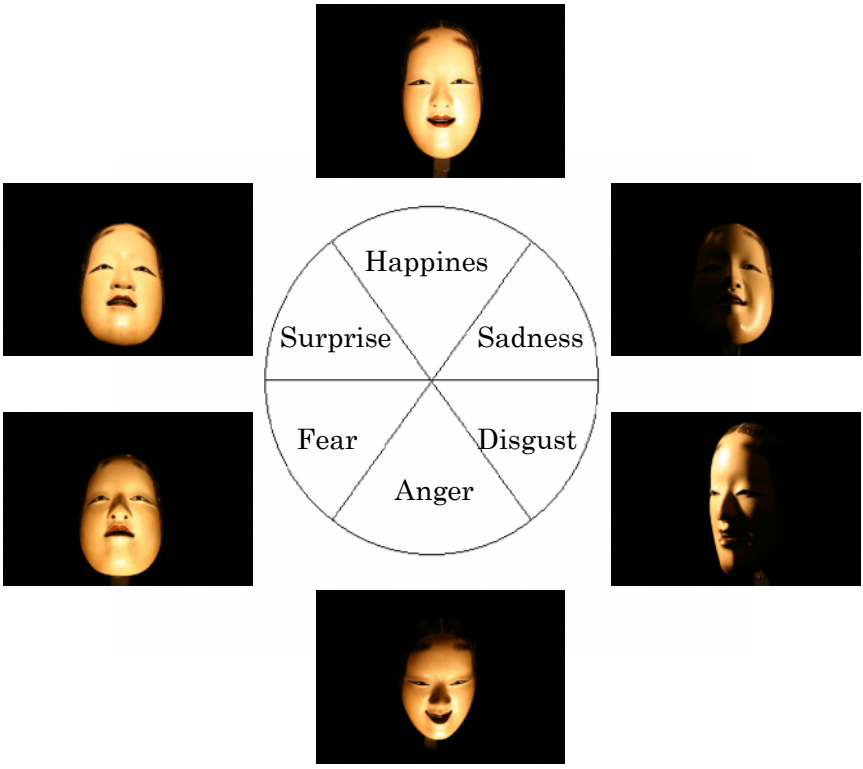


Figure 11.12 Emotion circle for the Noh mask

The examples of the sentences in the scenario [18] and the Noh masks selected by the system are shown in Figures 11.13–11.15. Figure 11.13 shows the scene where the angel is surprised to see a fisherman steal her cloak. Figure 11.14 shows the scene where the angel grieves and entreats the fisherman to give her cloak back. Figure 11.15 shows the scene where the angel is very glad to know that she will be able to go back to the heaven and she dances to show the glory of the heaven.

Angel
Stop!
That cloak is mine.
Why are you going with it?

Figure 11.13 Scene from *Hagoromo* (1)

Angel
Oh pitiful!
How shall I cloakless tread
the wing-ways of the air, how climb
the sky, my home?
Oh, give it back, in charity give it back.

Figure 11.14 Scene from *Hagoromo* (2)

Angel
I am happy, happy.
Now I shall have wings and mount the sky again.
And for thanksgiving I bequeath
a dance for remembrance to the world,
fit for the princes of men:
The dance tune that makes to turn
the towers of the moon,
I will dance it here and as an heirloom leave
it to the sorrowful men of the world.

Figure 11.15 Scene from *Hagoromo* (3)

11.3 Verbal Communication

In verbal communication, we can also make HCI pleasant. In this section, we explain our system which inserts "repeating the user's utterances" and "chiming in with the user" into the HCI [19].

When two persons converse joyfully, an unpleasant silence will fall upon them if one person does not reply to the other. However, chiming in with the speaker's remarks, or nodding, or repeating his/her utterances can encourage him/her to continue speaking, and can give pleasure to the communication, and can smooth the conversation.

This is also true of HCI [20]. If the computer cannot understand the user's responses and does not reply, the user will be discouraged and unwilling to continue the interaction. However, if the computer chimes in, or nods, or repeats his/her utterance, the user will be encouraged to continue the interaction. For these reasons, we have constructed human interfaces that use chiming in and repeating.

Figure 11.16 shows an example of the conversation on a "web-based system of health service" as follows:

System: Can you fill in the forms for your annuity by yourself?
User: I cannot always say that I cannot write it by myself.
System: Cannot you always say that you cannot write it by yourself?
User: I can fill it in a little.
System: Oh, really!

The user's first response to the question from the system is a double negative, and the system cannot understand whether the reply is affirmative or negative. However, in order not to stop the conversation and to encourage the user to continue, the system repeats the user's utterance in an interrogative sentence. Then, from the user's next reply "I can fill it in a little", the system can understand that the user can do it. At once, the system chimes in with him to show the hearer's interest.

If the system again repeats the user's utterance instead of chiming in, the user may go into details. In face-to-face communication, a double negative such as "I cannot always say that I cannot write it by myself" sometimes expresses a desire to say that the accident which prevented him/her from writing happened recently. For example, the system might be able to get the following reply:

System: Can you fill it in a little?
User: I hurt my right hand a few days ago, so I have difficulty in writing.

As shown in this example, repeating and chiming in to show the hearer's interest can make HCI pleasant.

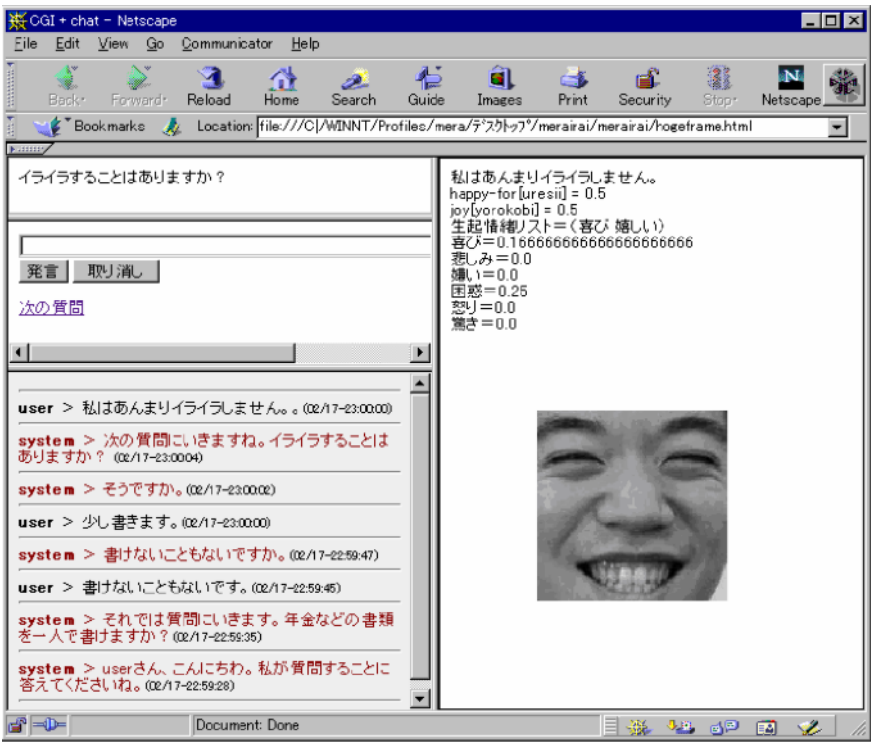


Figure 11.16 Interface for a "web-based system of health service"

11.4 Conclusions

We have tried to construct an emotion-oriented human interface. In this chapter, we have described several examples of our systems which enable human-like communication.

Both verbal and nonverbal messages are needed for smoothing human face-to-face communication. This is also true of HCI. In order to make HCI pleasant, we have adopted facial expressions as nonverbal messages and proposed fuzzy reasoning and neural network models for selecting the facial expression which expressed the emotions caused by several situations. Application of our models to e-mail shows that facial expressions can add impression of "pleasantness and familiarity" to the information on the computer display. Moreover, we applied our model to selecting and displaying a Noh mask in a Noh play scenario.

Chiming in with another's remarks, repeating their utterances, and so on, have the same functions as nonverbal messages. That is, chiming in and repeating can encourage a speaker to continue his/her speaking, give pleasure to the communication, and smooth the conversation. Since this is true of HCI, we have constructed human interfaces which chimed in with the user and repeated the his/her utterances.

However, many sciences such as physiology, psychology, medical science, philosophy have been struggling to understand human emotions. Our studies aim to achieve smooth and pleasant communications between humans and computers, but for further development of emotion-oriented HCI, a deeper understanding of human emotions is needed.

References

1. Ellison J.W., Massaro DW.: Featural evaluation, integration, and judgment of facial affect. J. Exp. Psychol. Human Perception Perform. **23**(1):213–226(1977)
2. Iwata M., Onisawa T.: Route decision system associating with facial expression. J. Jpn. Soc. Fuzzy Theor. Syst. **8**(3):532–540 (in Japanese) (1996)
3. Onisawa T., Date D.: Personified game playing support system. Jpn. J. Ergonomics **35**(3):157–167 (in Japanese) (1999)
4. Yamashita T. *et al.*: A study on characteristics of faces as a human interface. J. Jpn. Soc. Fuzzy Theor. Syst. **14**(2):248–253 (in Japanese) (2002)
5. Yamashita T.: An application of human mind to engineering. J. Soc. Biomech. **28**(2):72–78 (in Japanese) (2004)
6. Yamashita T., Koseki H.: An application of fuzzy reasoning to facial selection. Jpn. J. Edu. Technol. **22**(3):193–199 (in Japanese) (1998)
7. Yamashita T. *et al.*: An application of facial selection model by fuzzy reasoning to human interface. J. Jpn. Soc. Fuzzy Theor. Syst. **12**(2):313–320 (in Japanese) (2000)
8. Yamashita T., Sakai H.: Face selection model by fuzzy reasoning and its application to computer game. Jpn. J. Ergonomics **36**(3):139–145 (in Japanese) (2000)
9. Ichimura T. *et al.*: Approach to emotion oriented intelligent system by parallel sand glass type neural networks and emotion generating calculations. J. Human Interface Soc. **3**(4):225–238 (in Japanese) (2001)
10. Ichimura T. et al.: Emotion orientated intelligent systems. In: Howlett R.J. *et al.* (eds.) Internet-based Intelligent Information Processing Systems, pp. 183–226. World Scientific Publishing Company, Singapore (2003)
11. Plutchik R.: The multifactor-analytic theory of emotion. J. Psychol. **50**:153–171 (1960)
12. Matsumoto Y. *et al.*: Japanese Morphological Analysis System Chasen, Version 2.0. http://chasen-legacy.sourceforge.jp (in Japanese) (1999)
13. Yamashita T. *et al.*: Facial expressions as a human interface for e-mail. Mem. Tokyo Metropolitan Inst. Technol. **14**:7–11 (2000)
14. Osaka N.: Cross-cultural differences in the perception of facial expressions of ambiguous Noh faces. Bull. Psychonomic Soc. **24**(6):427–430 (1986)
15. Yamashita T. *et al.*: Recognition of affects in the facial expressions of Noh mask by fuzzy reasoning. Jpn. J. Ergonomics **35**(3):193–199 (in Japanese) (1999)
16. Yamashita T., Minoshita S.: Categorical perception in facial expressions on Noh mask "*koomote*." J. Jpn. Acad. Facial Stud. **5**(1): 49–57 (in Japanese) (2005)
17. Ekman P., Friesen V.W.: Unmasking the Face. Prentice-Hall, New Jersey (1975)
18. Waley A.: Hagoromo. http://etext.virginia.edu/etcbin/toccer-new2?id=WalHago.sgm&images=images/modeng&data=/texts/english/modeng/parsed&tag=public&part=1&division=div1. The University of Virginia Library Electronic Text Center (1997)
19. Sato Y. *et al.*: Proposal of automatic responding system for chat-system detecting affirmative/negative. In: Proceedings of 17th Fuzzy System Symposium, pp. 509–512 (2001)
20. Yamashita T.: Communication in terms of psychology: human computer interaction. Sci. Mach. **59**(1):198–203 (2007)