

EMOTIONAL SYNCHRONIZATION BASED HUMAN-ROBOT COMMUNICATION AND ITS EFFECTS

YILI

*Department of KANSEI Engineering
Division of Science and Technology, Shinshu University
3-15-1 Tokida, Ueda, Nagano 386-8567, Japan
liki@shinshu-u.ac.jp*

MINORU HASHIMOTO

*Department of Bioscience and Textile Technology
Interdisciplinary Graduate School of Science and Technology
Shinshu University
3-15-1 Tokida, Ueda, Nagano 386-8567, Japan
hashi@shinshu-u.ac.jp*

Received Day Month Year

Revised Day Month Year

Accepted Day Month Year

This paper presents a natural and comfortable communication system between human and robot based on synchronization to human emotional state using human facial expression recognition. The system consists of three parts: human emotion recognition, robotic emotion generation, and robotic emotion expression. The robot recognizes human emotion through human facial expressions, and robotic emotion is generated and synchronized with human emotion dynamically using a vector field of dynamics. The robot makes dynamically varying facial expressions to express its own emotions to the human. We conducted a communication experiment to examine the effectiveness of the proposed system. We found that subjects became much more comfortable after communicating with the robot with synchronized emotions. Subjects felt somewhat uncomfortable after communicating with the robot with non-synchronized emotions. During emotional synchronization, subjects communicated much more with the robot, and the communication time was double that during non-synchronization. Furthermore, in the case of emotional synchronization, subjects had good impressions of the robot, much better than the impressions in the case of non-synchronization. It was confirmed in this study that emotional synchronization in human-robot communication can be effective in making humans comfortable, and makes the robot much more favorable and acceptable to humans.

Keywords: Human-robot communication, facial expression recognition, emotional synchronization, vector field of dynamics.

1. Introduction

In the last few years, more and more robots have been created not only for the purpose of traditional industries but also for other fields such as medicine and education as well as daily living. We may use robots in the future for housekeeping, elderly care, and entertainment. Therefore, human-robot communication is a key issue. Natural and comfortable communication between humans and robots is attracting more and more attention. To naturally communicate, robots need to have not only intelligence but also affective behavior, such as emotions, feelings, impressions, sensitivity, and intuition. Humans communicate with one another through gaze direction, facial expression, body movement, speech, and language. To interact with humans in a human-like manner, robots must perceive and understand the richness and complexity of natural human social behavior. There are many studies on affective human-robot interaction that have explored how to make communication more natural and comfortable. For example, Kanda¹ reported the importance of cooperative behavior in a humanoid robot pretending to listen to route guidance from a human. In addition, effective gestures for human-robot interactions have been studied.^{2,3}

Facial expressions are of vital importance in expressing our emotions. It is the most intuitional and fast way to transfer our inner emotions to other people. Therefore, many facial expression robots have been created for human-robot communication, such as SAYA⁴ and Kismet.⁵ However, most of these robots are designed as mechanical structures. It is not easy to make rich facial expressions. We developed a facial robot KAMIN⁶ using a curved surface display as a face. This technique allows a facial expression to be easily made, compared with other mechanical methods. We proposed how to generate a robot's facial expressions based on the characteristic quantity in its emotional space.⁷

In addition to facial expressions, emotional synchronization is an important concept in human communication. There are many phenomena of emotional synchronization that happen in our daily lives. Indeed, some synchronized phenomena are well-known and have been discussed related to human interaction.^{8,9} Lundqvist¹⁰ found that synchronized reactions occur in happy, angry, and sad expressions during human communication when partners change their expressions. Ichikawa et al.¹¹ reported that happy expressions are strengthened, whereas angry and sad expressions are weakened by synchronization. Jonsson et al.¹² studied the influence of a car voice on a driver's behavior and attitudes. The participants' emotion was first induced by video clips. Then participants used a driving simulator and interacted with the car voice. They found that drivers who interacted with voices that matched their own emotional state had less than half as many accidents on average as drivers who interacted with mismatched voices. Furthermore, drivers paired with matched voices also communicated much more with the voice. Our previous study¹³ included experiments to evaluate the effects of emotional synchronization in human-robot communications using voice recognition. In the experiments, human emotion was recognized by the voice analysis software RobEsense. The robot KAMIN expressed its emotion by making different facial expressions. We found that subjects became more comfortable when the robot made synchronized facial expressions in response to human emotion. Therefore, emotional synchronization is an important issue and often has positive effects not only on human communication, but also on human-robot communication.

Researchers have proposed utilizing nonverbal mirroring of learners' behaviors including facial expressions to provide affective support to them, with the effect of a mimicking agent on a learners' performance investigated.^{14,15} Riek et al.¹⁶ proposed a mimicking behavior of a human head gesture to enable human-robot rapport. They took methods to mimic only behaviors or facial expressions without concerning emotion. However, it is important to make emotional synchronization with the partner's emotion for healing sadness or loneliness likely in human-human communications. There are relationships between expressions such as facial expression, body behavior, and voice at the same emotional state. If we can make synchronization in emotional state, we can express using some modalities at the emotional state.

The emotion is a dynamical system not a static one³⁵. The emotional state change dynamically. In order to change emotion naturally, it is important to make a dynamical system of robotic emotion. The former researches for mimicking interaction made just same expressions with the partner's expressions. This mimicking method is a static system. Though mirroring can attract and give a friendship, it's difficult to touch a person's heart by sharing emotion each other.

In our study, we propose a system for human-robot communication to synchronize the robot emotion with the human emotion to cause sympathy. Then we investigate how human emotions and impressions are affected by a robot that has emotional entrainment ability. Emotional recognition was implemented by processing facial expressions, and emotional entrainment was realized based on a dynamic vector field. We think that it can generate natural and continuous changing of the facial expression by entrainment in the vector field of emotion because the facial expression changes dynamically and continuously. In the case of mimicking, it's just static. Therefore, our approach is different from the aforementioned mirroring technique because it is based on entrainment of emotions between human and robot, not behavioral mimicry.

Many researchers¹⁷⁻²⁰ have investigated emotion-generating methods for human-robot interactions. Kim et al.¹⁷ introduced a computer model of emotion generation based on cognitive appraisal theory. Interactive emotion recognition using a support vector machine has been conducted¹⁹. Kirby et al.²⁰ proposed an affective model based on the use of expressive moods and emotions to realize human-robot interaction proceeding in a smooth, natural manner. However, those works did not involve emotional synchronization.

In this study, we focused on communication between human and robot. To make natural and human-like communication, we proposed a communication system based on emotional synchronization using facial expression recognition in human-robot communication. The robot recognizes human emotion through human facial expressions. Robotic emotion is synchronized with human emotion dynamically using a vector field of dynamics. Then, the robot makes a facial expression to express its own emotion. We conducted experiments to evaluate the effectiveness of the proposed system by looking at how human emotion changes through interacting with a robot. The potential effect of emotional entrainment could be that a happy reaction induces a happier response compared to the original emotion expressed. A sad reaction could weaken the response of sadness compared to the original emotion expressed. Anger could be an exception, with an angry reaction escalating emotions and worsening the interaction.

In the next section, we address the outline of the communication system. In section III, we discuss how to analyze human facial expressions. In section IV, we explain the emotional synchronization system and show the simulation results. In section V, we present the communication experiment. We show the main experimental results in section VI and offer conclusions in section VII.

2. Overview of the Communication System

The communication system consists of three parts: face recognition, robotic emotion generation, and robotic emotion expression, as shown in Fig. 1. The robot recognizes human emotion X_H through human facial expression, and robotic emotion X_R is created and synchronized with human emotion using the vector field of dynamics. The robot recognizes human emotion dynamically, and the robotic emotion is entrained to human emotion in the vector field of dynamics. In that way, the robot can have a dynamic emotional change like humans. In addition, it makes possible real-time communication between human and robot and also makes communication natural and human-like. In the emotion expression part, the robot expresses its emotion by facial expressions. For example, when the robot recognizes that the human is in a happy mood, it also feels happy and expresses it with a happy facial expression. This process is continuous, with communication between human and robot based on emotional synchronization.

We established the relation of human emotion recognition and synchronization to robotic emotion expression using four spaces: emotion recognition space, emotion generation space, symbol space, and emotion expression space. Figure 2 shows the relationships among the four spaces. First, the human emotional state is recognized in the emotion recognition space. Second, the human emotional state is mapped into the emotion generation space, and the robot recognizes it and synchronizes its own emotion with it using the vector field of dynamics. The synchronized emotion of the robot is mapped into the symbol space to figure out the corresponding state vector for the synchronized robotic emotion. Finally, the determined state vector in the symbol space is mapped to the emotion expression space, and the robot expresses its synchronized emotion to the human using robotic facial expressions dynamically in the emotion expression space. Then the human changes emotion. By iterating the second to fourth steps, the robot communicates with the human continuously. Thus, real-time natural communication based on emotional synchronization is realized.

3. Face Recognition

Facial analysis is one of the key parameters for emotion recognition in human-robot interaction. It provides a natural and efficient way to communicate between human and robot. Much information about a person's emotions and state of mind can be obtained from their facial expressions alone. Indeed, research in psychology^{21,22} has shown that facial expressions play a major role in human conversation coordination and have a greater influence on auditors than the textual content of a spoken message.

A face recognition system generally consists of image acquisition, face detection and tracking, facial feature extraction, and emotional classification. Various approaches to face detection, facial feature extraction, and facial expression recognition have been reported in the literature over the last few decades such as the eigenface based on principal component analysis,²³ geometric modeling,²⁴ deformable template,²⁵ neural networks,²⁶ and color analysis.²⁷ However, most computer vision-based approaches to facial expression analysis so far have somewhat complicated algorithms with a large amount of calculations attempting to recognize only a small set of prototypic expressions of emotion (i.e. happiness, surprise, anger, sadness, fear, and disgust). However, in human communication, we not only communicate with others with emotions in terms of categories (happiness, anger, fear, etc.) but also with emotions that span the relationship between different emotions (e.g. an elated emotion blends excited with joyful). Therefore, to make natural and human-like communication between human and robot, it is very important to consider such issues, although few studies have addressed them so far. Russell²⁸ proposed a circumplex model of affect, widely recognized and used, and argued that the human observer perceives two broad affective categories on the face: arousal and pleasantness. Yamada et al.^{29,30} proposed that a relationship exists between the relative displacement of facial expressions and basic affect categories, and can be indicated using two variables: curving and openness, and inclination.

Generally speaking, facial expressions are the visual changes in the face due to the actions of facial muscles. Therefore, it is reasonable to describe facial expressions in terms of the changes in appearance of the face. In this study, to enhance efficiency, we used the least facial features for recognition of human emotion. We referred to Yamada's theory to design a 2D continuous physical space to represent the emotional states for recognizing

human emotion. The algorithm in this study has many advantages. First, it has the lowest dimension of the space-based recognition methods. Thus, recognition should be easier. Second, because it uses the least features to make a space, it reduces the amount of calculation and saves time, making it possible to establish a real-time communication system. Third, it can be implemented easily within a short period. Furthermore, because of the continuous space of emotion, it is possible to recognize not only a small set of prototypic expressions of emotions, but also additional ones (e.g. elated and relaxed), making the robot much more capable of understanding humans and enriching the emotion and expression in itself.

3.1. Face detection and tracking

In this study, we used the AdaBoost algorithm-based face detector³¹ that employs Haar-like classifiers arranged in a cascade structure, with high accuracy and robustness against observations with low resolutions or varying illumination conditions. We applied the AdaBoost algorithm-based face detector to our system to detect the face and facial features (brow, eye, and mouth). In our system, the face detector can run at a rapid speed of 25 frames per second with accuracy over 95%, making it possible to make real-time human-robot communication.

3.2. Facial feature extraction

Facial feature extraction refers to the capability of identifying facial images seen by a robot in detail. After detection and tracking of facial features such as the eye and mouth regions, we preprocessed each region with contrast and binarization, then divided the eye region into brow and eye and processed them again to make them much clearer and closer to the shape of the real eye and eyebrow on the face. Furthermore, we subdivided the brow and eye into left and right parts because the human face is not bilaterally symmetrical. As for the mouth, the color of the mouth is somewhat similar to the color of the face skin. It is not easy to extract it from the background using just contrast and binarization. Therefore, we converted the RGB (red, green, blue) color system to the HSV (hue, saturation, value) color system for the mouth region. The HSV color system has many advantages. For example, the hue and saturation of the space are insensitive to brightness, and this system has a better sense of color, making it easy to do the analysis. Figure 3 shows processed results of the brow, eye, and mouth online. Finally, we scanned the black pixel in each divided region to locate the exact position and get the details of each feature such as length, width, and angle (see Fig. 4).

3.3. Facial expression recognition

3.3.1. Yamada's theory

Yamada et al.^{29,30} attempted to identify a model concerning perceptual judgments of emotion from facial expressions. They found a relationship between basic emotional categories and structural variables of facial expression based on the displacement structure of the characteristic points in a facial expression. They proposed that the categorical judgments of emotions could be explained well by two canonical variables: bending and inclination. Bending involves displacement of feature points related to the amount of eyebrow curving and eye and mouth opening. Inclination involves displacement of the feature point concerned with the angles of the eyes and eyebrows, and the extent of the V or inverted V-formation of the mouth (see Fig. 5).

3.3.2. Definition of bending and inclination

As shown in the left part of Fig. 6, we drew a line from the outer brow point to the inner brow point and defined the inclination as the tangent of the angle between that line and a horizontal line. We drew two lines from the outer lip corner point and inner lip corner point to the middle lips point, and defined the inclination of the mouth as the average of the tangents of the included angle of both sides. We defined bending as the quotient of the height and length of each part, as shown in the right part of Fig. 6. We defined the inclination and bending for brow and eye as the average value of the left and right parts, and defined the inclination for mouth as the average value of the left and right parts.

In addition, we considered the state of "normal" as the reference base, and defined the signs of inclination and bending as shown in Fig. 7. The inclination of each facial expression was defined as the sum of inclination of brow, eye, and mouth, and the bending of each facial expression was defined as the sum of bending of brow, eye, and mouth. Therefore, the practical inclination (**I**) and bending (**B**) of each facial expression can be

computed using the following equations:

$$\mathbf{I}_i = \sum_{j:\text{brow,eye,mouth}} \tan \Theta_{j,i} - \sum_{j:\text{brow,eye,mouth}} \tan \Theta_{j,\text{normal}} \quad (1)$$

$$\mathbf{B}_i = \sum_{j:\text{brow,eye,mouth}} \frac{H_{j,i}}{L_{j,i}} - \sum_{j:\text{brow,eye,mouth}} \frac{H_{j,\text{normal}}}{L_{j,\text{normal}}} \quad (2)$$

3.3.3. Emotion recognition space

After defining inclination and bending for the facial expressions, we conducted an experiment to learn the 2D physical space. We used 43 subjects' six basic expressions (happiness, surprise, fear, anger, disgust, and sadness) from three databases, as 10 subjects of the Japanese Female Facial Expression (JAFPE) database,³⁶ 32 subjects of the Bosphorus database³⁷ and the author's six basic expressions database. We preprocessed the images (with luminance, contrast, size, and position) and extracted the facial features of each expression by hand to get the inclination and bending. Because the facial expressions of 'happiness' and 'anger' had the maximum and minimum inclination, respectively, and the facial expression of 'surprise' and 'sadness' had the maximum and minimum bending, respectively, we used 'happiness' and 'anger' to normalize the inclination as 1 and -1 and used 'surprise' and 'sadness' to normalize the bending as 1 and -1 respectively. Figure 8 shows the normalized 2D physical space for the six basic expressions.

3.4. Experiment of facial expression recognition

To examine the effectiveness of the facial expression recognition system, we conducted an experiment to check the recognition rate. We asked 12 college students (6 males and 6 females of Japanese and Chinese, with a average age of 23) to make the six basic expressions (happiness, surprise, fear, anger, sadness, and disgust) at random in front of the digital camera. To make the subjects' facial expressions natural and consistent with their emotional state, no sample expression images were exposed to the subjects beforehand. When making each expression, each subject was induced by the same suggestions, pictures, and videos. For example, to make the expression of happiness, we suggested that the subject think about something that made them feel happy and also showed them some pictures portraying comfort and happiness. The same method was used for the other five expressions. Figure 9 shows an overview of the experiment. The subjects sat about 50 cm in front of the camera and were asked to maintain the position of their heads when making the expressions. The parameter of inclination and bending of the expression were extracted automatically by the image processing system online. Results were mapped into the emotion recognition space. Then, we computed the closest distances between the point we mapped and the points distributed in the space to find out the emotional state of the current facial expression. For each expression, we counted the number of frames that hit the right expression within about 40 frames. We obtained an average recognition rate of 77% for the six basic facial expressions of the 12 subjects. Figure 10 shows the average distribution of the 12 subjects' six basic expressions. We can see that the distribution is the same as the distribution of the 2D emotion recognition space we obtained, indicating that our online facial expression recognition system was capable of recognizing human emotion automatically, making it possible to realize human-robot communication in the following procedures.

4. Emotional Synchronization and Expression

4.1. KAMIN and Its affect model

In this study, for Human-Robot communication, we used a head robot KAMIN⁶ as shown in Fig. 11 (a). The head mechanism is a facial image display, and it consists of a dome screen, a fish-eye lens, and a projector. The face image is projected to the dome screen from the inside. The fish-eye lens is installed on the front of the projector, and it projects the facial expressions on the dome screen. The neck movement is also possible with four degrees of freedom by using four motors. By using this head robot we can make various facial expressions easily compared with methods of mechanical facial expression. Also there is a three-dimensional effect caused by the curved surface, which is not in a plane image. Furthermore, in human communication, the movement of a head affects the impression of face expression. So more impressive face expression is also possible by this robot when a facial image display and a head movement mechanism are integrated and cooperate.

We assumed that a facial expression is not static, it is a dynamic process (e.g. when we are happy, we will naturally have a happy expression, although the happy expression is not the same in appearance as time goes by). That is, even when the robotic emotion does not change, the facial expression of the emotion changes dynamically to a slight degree. Thus, the robot is much more attractive and lively, making itself more compatible with humans. Our previous study⁷ proposed making dynamical robotic facial expressions by changing the characteristic quantity in a 2D space defined by inclination and bending (see Fig. 11 (b)), using the vector field of dynamics.

If we assume the robot has its own emotional state, it is necessary to consider the affect model of the robot. In this study, the robotic affect model was constructed based on a human affect model. From the numerous human affect models proposed, we chose the circumplex model (see Fig. 12) proposed by Russell²⁸ because it is a simple 2D model, it is easy to construct the system with the same dimensional 2D facial expression space of the robot. And instead of viewing emotions in terms of categories (happiness, anger, fear, etc.), this viewpoint of the affect model conceptualizes the dimensions that can span the relationship between different emotions. This makes a lot of sense for human-robot communication if we apply it to the robotic affect model.

4.2. Dynamic-based information processing

We applied dynamic-based information processing³² to make emotional synchronization in human-robot communication. In this section, we will discuss the relationship between dynamic and robotic emotional synchronization.

Consider human emotion E . The time sequence data of this motion is assumed to be $\zeta[k]$ (e.g. happiness, anger) that composes E as follows:

$$E = [\zeta[1] \quad \zeta[2] \quad \cdots \quad \zeta[m]] \quad (3)$$

$$\zeta[k] = [\xi_1[k] \quad \xi_2[k] \quad \cdots \quad \xi_N[k]]^T \quad (4)$$

where m means a number of data and N means the number of parameters (or dimensions) of human emotion. Because E composes a curved line in N dimensional space, when E composes a cyclic emotion, E shows the closed curved line.

On the other hand, consider the dynamics represented by the following difference equation:

$$\mathbf{x}[k+1] = \mathbf{x}[k] + f(\mathbf{x}[k]) \quad (5)$$

$$\mathbf{x}[k] = [x_1[k] \quad x_2[k] \quad \cdots \quad x_N[k]]^T \quad (6)$$

Suppose that E is an attractor of these dynamics, the state vector $\mathbf{x}[k]$ starts from the initial value $\mathbf{x}[0]$ and converges to the following equation:

$$\lim_{k \rightarrow \infty} \mathbf{x}[k] = \xi[k + k_0] \quad (7)$$

where k_0 depends on x_0 .

In this case, the dynamics memorize and reproduce the whole emotion E . By picking up $\mathbf{x}[k]$ ($k=0, 1, 2, \dots$), we can obtain $\xi[k]$ ($k=0, 1, 2, \dots$), which means that the dynamics reproduce time sequence data E of emotion.

4.2.1. Design algorithm of the dynamics

As in (5), we can suppose that $\mathbf{f}(\mathbf{x}[k])$ defines the vector field in N dimensional space. By using the polynomial functional approximation of the vector field, the dynamics in (5) can be calculated. The algorithm design of the dynamics is as follows.

Step 1: Draw the closed curved line E in N dimensional space.

Step 2: Set the basin D of attractor and define sample points η_i and vector of $\mathbf{f}(\eta_i)$ making the closed curved line E be an attractor. Figure 13 shows the definition of the vector field.

The vector field is formed around arbitrary curve E , which is supposed as an attractor, and region D is a basin of entrainment around the attractor. Curve E is defined as in section 4.1.1. The number of sample points can be decided among the basin of entrainment around the attractor which is described as η_i ($i = 1, 2, \dots, m$),

where m is the number of sample points. $\xi^{\eta_i}[k]$, located on the attractor, is the nearest point from η_i , and $\delta_i[k]$ is the connection vector between η_i and $\delta_i[k]$. Then, $\delta_i[k+1]$ and $\delta_i[k]$ can be defined as follows:

$$\delta_i[k+1] = (\eta_i + f(\eta_i)) - \xi^{\eta_i}[k+1] \quad (8)$$

$$\delta_i[k] = \eta_i - \xi^{\eta_i}[k] \quad (9)$$

The sufficient condition for convergence is

$$\|\delta_i[k+1]\| < \|\delta_i[k]\| \quad (10)$$

Step 3: The defined vector $f(\eta_i)$ is approximated by the following equation by e -the order polynomial of $\mathbf{x}[k]$.

$$f(\eta) = \sum_{p=0}^e \sum_{p_1, \dots, p_N}^e a_{(p_1 p_2 \dots p_N)} \prod_{i=1}^N \eta_i^{p_i} \quad (11)$$

$$\sum_{p_i: \text{positive integer}} p_i = P \quad (12)$$

$$\eta = [\eta_1 \quad \eta_2 \quad \dots \quad \eta_N]^T \quad (13)$$

Where $a_{(p_1 p_2 \dots p_N)}$ is a constant.

Here, defining $f(\eta)$ as:

$$f(\eta) = \Phi(a_{(p_1 p_2 \dots p_N)}) \theta(\eta) \quad (14)$$

$$\theta(\eta) = [\eta_1^e \quad \dots \quad \eta_N^e \quad \eta_1^{e-1} \eta_2 \quad \dots \quad 1]^T \quad (15)$$

Φ is calculated by the least squares method as follows.

$$\Phi(a_{(p_1 p_2 \dots p_N)}) = F \Theta^\# \quad (16)$$

here $\Theta^\#$ means the pseudo inverse matrix of Θ .

$$F = [f(\eta_1) \quad f(\eta_2) \quad \dots \quad f(\eta_m)] \quad (17)$$

$$\Theta = [\theta(\eta_1) \quad \theta(\eta_2) \quad \dots \quad \theta(\eta_m)] \quad (18)$$

Φ is a constant parameter matrix that defines the dynamics in (5). In this way, as long as we can well approximate to the vector field defined by Φ , any point $\mathbf{x}[0]$ from the basin D can entrain to the attractor E when $k \rightarrow \infty$.

4.2.2. Online design of the dynamics

In (16), Φ which defines $f(\mathbf{x}[k])$ is designed based on the least squares method. Online design of the dynamics is used to design the online least squares method, meaning the dynamics memorize the human emotion successively. The parameter matrix Φ_m in time sequence m is calculated by the iteration of the following online least squares algorithm using a non-singular matrix P_m .

$$P_{m+1} = P_m - \frac{P_m \theta(\eta_{m+1}) \theta^T(\eta_{m+1}) P_m}{1 + \theta^T(\eta_{m+1}) P_m \theta(\eta_{m+1})} \quad (19)$$

$$\Phi_{m+1} = \Phi_m - \{ \Phi_m \theta(\eta_{m+1}) - f(\eta_{m+1}) \} \theta^T(\eta_{m+1}) P_{m+1} \quad (20)$$

P_m is defined as

$$P_m = (\Theta\Theta^T)^{-1} \quad (21)$$

where Θ is defined in (18).

By using (19) and (20), the dynamics defined by equation (5) that has an attractor can be gradually designed online.

4.2.3. Robotic emotion generation and synchronization

We designed the vector field of dynamics on Russell's 2D space to realize the entrainment between human and robotic emotions. We call it the emotional generation space for the robot as shown in Fig. 14. The recognized human emotion is mapped into the vector field as one point, and an attractor (human emotion) is constructed in the vector field using the online design method of dynamics. The attractor is updated continuously according to the result of the recognition part. Then the robotic emotion is entrained to human emotion dynamically in the vector field. The emotional synchronization is realized.

We find an interesting relation between the 2D physical space and Russell's emotional space. We can see that the distribution of the 2D physical space is quite similar to the middle and upper part of Russell's space except the origins are different (we suppose the position of 'disgust' in the emotion generation space is near 'angry' since they have the similar semantic meaning). If we pay attention only to the relationship between 'surprise' and 'sad' in the Y axis direction and the 'happy' and 'angry' in the X axis direction on the 2D physical space and Russell's space. And Yamada also discussed that the dimensions of visual information defined as inclination and bending are corresponding to those of semantic affective meanings found in earlier research as pleasantness and activity, respectively.³⁸ It seems to suggest that the dimensions of the 2D physical space are corresponding to the dimensions of affective semantic meanings, i.e. inclination and bending of the 2D physical space are corresponding to the levels of pleasantness and arousal of Russell's emotional space respectively. And then we get the affective meaning for the 2D physical space to represent human emotional states. So in such view, it is reasonable to use the facial expressions to map on the space to recognize human emotions. So in this study we translate the emotion recognition space to the emotion generation space as shown in Fig. 15 to map on the recognized human emotion. The basic emotions transferred to the results of the facial expression recognition experiment as seen in Fig. 10.

4.3. Robotic emotion expression

After recognizing and synchronizing with human emotion, the robot KAMIN expresses its synchronized emotion by its own facial expressions. We proposed to make the robotic facial expressions by changing the characteristic quantity in a 2D space defined by inclination and bending, using the vector field of dynamics. That is, the facial expression of the robotic emotion changes dynamically among different emotions which makes the robot humanoid. The change in inclination involves the displacement of the feature point concerned with the angles of the eyes and eyebrows, and the extent of the V or inverted V formation of the mouth. The change in bending involves the displacement of feature points related to the amount of eyebrow curving and eye and mouth opening. Figure 16 shows an example of five expressions for robotic emotional state (normal, happiness, anger, sadness, and surprise).

4.4. Symbolization

Because robots, such as a humanoid robot, can have many emotions like humans, we designed the robotic facial expressions as a dynamical process even for the same emotion. However, the vector field of emotion increases, and designing the dynamics takes a longer time. To enhance the efficiency of the information processing in real time, it is necessary to design the vector field based dynamics in the continuous symbol space. In this study, we referred to Okada's method³³ to symbolize the robotic expression space to a continuous symbol space. Figure 17 shows the concept of the continuous symbol space. One point in the symbol space defines one dynamic in the emotion expression space. The state vector moves following the vector field of the dynamics in the emotion expression space, and the state vector in the symbol space moves following the dynamics, which changes the configuration of the dynamics in the emotion expression space. In this way, the emotion expression and transition of the robot is realized.

4.5. Non-linear mapping

Nonlinear mapping is a method often used to establish connection between two spaces, especially those with different dimensions. In this study, we designed nonlinear mapping based on the polynomial functional approximation³⁴ to connect the emotion generation space to the symbol space of the robotic emotion expression space. Figure 18 shows the concept of the relation between the emotion generation space and the symbol space. One state vector in the emotion generation space corresponds to one state vector in the symbol space.

4.6. Simulation of the communication system

We did a simulation of the emotional synchronization-based communication system. We supposed that human emotion changes in an order of ‘normal-happiness-surprise-anger-sadness-normal’. The corresponding changes in the emotion generation space, symbol space, and emotion expression space are shown in Fig. 19. Figure 20 shows some corresponding facial expressions of the states marked in the expression space. The robot recognizes each emotion dynamically and expresses the synchronized emotion dynamically in the emotion expression space. The results of the simulation suggested that the dynamic-based information processing system worked effectively in the proposed system.

5. Communication Experiments

5.1. Purpose and hypothesis

To evaluate the effectiveness of the proposed system based on emotional synchronization, we have to examine the influence of emotional synchronization in communication between human and robot. We had three hypotheses: 1) the emotional synchronization-based communication would make humans feel comfortable, 2) humans would be willing to communicate with such a robot, and 3) humans would have positive impressions and become more accepting of a robot that can communicate by emotional synchronization.

5.2. Method and condition

We conducted a communication experiment between human and robot to examine the influence of emotional synchronization on human emotional state during human-robot communication. To evaluate the effect of synchronization, we designed synchronization and non-synchronization cases. In the synchronization case, the robotic emotional state was synchronized with the human emotional state (see Fig. 21). In the non-synchronization case, we designed it so the robotic emotion was the very reverse of the human emotional state (see Fig. 22). That is, if the human emotional state was pleasant or unpleasant, the robotic emotional state became reversely unpleasant or unpleasant. If the human emotional state was arousal or non-arousal, the robotic emotional state reversely became non-arousal or arousal. This series of experiments was designed to investigate the effect of emotion entrainment. Therefore, we compared different effects by implementing different emotions, one being synchronized emotion and the other the opposite emotion (non-synchronized) in the communication robot. If we gave the communication robot no emotion, we thought it highly likely that the participant’s impression would be determined based on whether there was an emotional reaction. Therefore, to investigate the effect of emotion synchronization, it was necessary to provide a reaction emotion according to the participant’s emotion. In this study, we provided the opposite emotion to the participant’s emotion as the object for comparison.

The methodology of the communication experiment was as follows.

- (a) Forty university students (20 males and 20 females) were asked to take part in the communication experiment. They were not familiar with robots.
- (b) The 40 subjects were divided into two groups. One group (20 subjects, 10 males and 10 females) was only asked to participate in the synchronization communication. The other group (20 subjects, 10 males and 10 females) was only asked to participate in the non-synchronization communication.
- (c) Subjects sat about 50 cm in front of the robot KAMIN (see Fig. 23). The camera was set in front of the robot. Subjects were asked to communicate with the robot freely, just using their facial expressions to reflect their internal emotional states. That is, they expressed facial expressions consistent with their emotional state (e.g. subjects had happy facial expressions when they felt happy). They were asked not to make large head movements during the communication.

- (d) The length of the communication experiment was determined by each subject, with each saying 'OK' as a cue for ending the communication.
- (e) Subjects were asked to complete questionnaires before and after the experiment.
- (f) Subjects were not given information about synchronization or non-synchronization before the experiment.

The procedure of the communication experiment was as follows.

Step 1: Before the communication experiment, subjects were asked to complete a questionnaire about their emotional state based on two items: 'discomfort-comfort' and 'sleepy-aroused' in five degrees from -2 to 2.

Step 2: Subjects were asked to sit in front of the robot and adjust the camera to fit correctly for them.

Step 3: We modified the image processing system to fit the feature extraction for subjects by asking them to make some basic expressions (happiness, anger, surprise, etc.). Subjects communicated with the robot using their natural facial expressions. However, because some people have difficulty expressing emotions (as learned from the recognition experiment), to enhance the recognition accuracy for subjects, we rescaled the recognition space for each subject in this step.

Step 4: Subjects were asked to communicate with the robot Kamin freely on a verbal cue of 'Start'.

Step 5: After the communication experiment, subjects were asked to complete the same questionnaire as in step 1.

Step 6: Subjects were asked to fill out an impression evaluation questionnaire for the robot Kamin using a semantic differential (SD) method with a 5-point scale, one of the most widely-used scales for the measurement of attitudes. The items of the questionnaire were determined by free discussion among researchers concerning communication robots.

The semantic differential items were determined based on free discussion among the researchers about their impressions of a communication robot. The adjective items and their opposites from the discussion were used as the semantic differential items. A 5-point scale was used to measure participants' impressions.

6. Results and Discussion

In the synchronization case, the robotic emotion was synchronized with the recognized human emotion, that is, when humans felt pleasant or unpleasant, the robot also felt pleasant or unpleasant. When human emotion was arousal or non-arousal, the robotic emotion became arousal or non-arousal. In non-synchronization, the robotic emotion was the reverse of the human emotion. This indicates that the robot was capable of recognizing human emotion correctly from facial expressions and responded properly, according to the human emotional state. This confirmed that the image processing system and dynamic-based information processing system worked effectively. Figures 24 and 25 show the average results of the questionnaires for the 40 subjects. In the case of synchronization, both comfort and arousal of human emotional state changed greatly after the experiment with a significant difference at the 1% level. We conducted t-tests and calculated the p-values and effect sizes. The p-values between before and after human/robot interaction in the case of synchronization were 3.0E-06 (comfort) and 4.5 E-06 (arousal), and effect sizes were 1.26 (comfort) and 1.31 (arousal). This suggests that emotional synchronization to human emotional state in human-robot communication can facilitate a comfortable state in humans and arouse enthusiasm. Subjects may have felt more interested and excited after the emotional synchronization-based communication with the robot. In the case of non-synchronization, although the average value for comfort became lower after the experiment, arousal became higher with a significant difference at the 1% level. The p-value for arousal in the case of non-synchronization was 6.4E-05, and the effect size was 1.03. These results suggest that non-synchronization of human emotional state in human-robot communication might lead to an uncomfortable state in humans. Increased arousal might indicate that uncomfortable communication makes humans feel tense and upset or just awakens them from a drowsy state. We can thus conclude that the emotional synchronization of human emotional state in human-robot communication had a positive influence on human emotional state. During emotional synchronization, subjects communicated much more with the robot, with communication time double that during non-synchronization (see Fig. 26). Furthermore, in the case of emotional synchronization, subjects had good impressions of the robot, much better than those in the case of non-synchronization (see Fig. 27). This suggests that humans are much more willing to communicate with and perhaps are more accepting of a robot that can synchronize with their emotions.

7. Conclusions

This paper proposed a communication system based on emotional synchronization to human emotion using

facial expression recognition in human-robot communication. The robot recognized human emotion through human facial expressions. Robotic emotion was synchronized with human emotion using a vector field of dynamics, and the robot made facial expressions to express its own emotions.

We proposed a method using the AdaBoost-based face detector to detect and track human facial features and developed a method to get the facial features in real-time using an image processing technique. The detector we developed can run at a speed of 25 frames per second, which allows enough time to do the following process. Based on Yamada's theory, we used facial features to make a 2D continuous physical space defined by two parameters, inclination and bending, to represent human emotional state to recognize human emotions. This facial expression recognition system has the merits of performing at a rapid speed and recognizing a large set of emotional states, making it possible to use it in real-time communication between human and robot. In emotional synchronization, we proposed an approach using a dynamic-based information processing system to generate robotic emotion and synchronize it with recognized human emotion dynamically using a vector field. The robot could dynamically vary emotional states and generate its own emotion, synchronizing it with human emotion dynamically to ensure a timely and continuous reaction to humans during communication.

In this study, we conducted experiments to evaluate the effectiveness of the proposed system based on emotional synchronization. We found that subjects became much more comfortable after emotionally synchronized communication with the robot. They communicated much more with the robot and had better impressions of it compared to non-synchronized communication. We confirmed that emotional synchronization in human-robot communication can be effective in making humans comfortable and making a robot much more favorable and acceptable to humans.

In the future work, in order to make the communication much more natural, comfortable and intelligent between human and robot, we intend to improve the system to be capable to detect and express more different emotional queues, e.g. recognize and express more natural facial expressions with head motion or use more complex realistic facial expressions for the robot. And to make the degree of synchronization adjustable according to the human emotional state during the communication.

We also plan to make an algorithm to integrate the facial expression recognition algorithm, voice recognition algorithm and head motion algorithm into the system which can more effectively make the human-like comfortable communication between human and robot.

References

1. Takayuki Kanda, Masayuki Kamasima, Michita Imai et al., A humanoid robot that pretends to listen to route guidance from a human, *Autonomous Robots*, **22**(1)(2007) 87-100.
2. Candace L. Sidner, Christopher Lee, Louis-Philippe Morency, Clifton Forlines, The effect of head nod recognition in human-robot conversation, in *ACM Conf. on Human-Robot Interaction (HRI)* (ACM Press, Salt Lake City, Utah, USA, 2006), pp.290-296.
3. C. Pelachaud, Studies on gesture expressivity for a virtual agent, *Speech Communication*, **51**(7)(2009) 630-639.
4. Takuya Hashimoto, Sachio Hiramatsu, Toshiaki Tsuji and Hiroshi Kobayashi, Development of the Face Robot SAYA for Rich Facial Expressions, in *SICE-ICASE Int. Joint Conf.* (SICE-ICASE Press, Busan, Korea, 2006), pp. 5423-5428.
5. <http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>.
6. M. Hashimoto, D. Morooka, Robotic Facial Expression Using a Curved Surface Display, *Robotics and Mechatronics*, **18**(4)(2006)504-510.
7. T. Usui, K. Kume, M. Yamano and M. Hashimoto, A Robotic KANSEI Communication System Based on Emotional Synchronization, in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)* (IEEE Press, Acropolis Convention Center Nice, France, 2008), pp. 22-26.
8. Hinsz V.B., Tomhave J., Smile and (half) the world smiles with you, frown and you frown alone, *Personality and Social Psychology Bulletin*, **17**(1991) 586-592.
9. Hess U., Philippot P. and Blairy S., Facial reactions to emotional facial expressions: Affect or cognition?, *Cognition and Emotion*, **12**(1998)509-531.
10. Lundqvist L.O., Dimberg U., Facial expressions are contagious, *Psychology*, **9**(3)(1995)203-211.
11. Hiroko ICHIKAWA, N.P.CHANSRASIRI, Hiroshi HARASHIMA and Junshiro MAKINO, The effect of congruent facial expression on counterpart's facial response, in *Technical report of IEICE, PRMU2002-210*, **102**(651)(2003), pp. 59-64. [in Japanese]
12. I.M. Jonsson, C. Nass, H. Harris, L. Takayama, Matching In-Car Voice with Driver State Impact on Attitude and Driving Performance, in *Proc. of the Third Informational Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (Rockport, USA, 2005), pp. 173-180.
13. M. Hashimoto, M. Yamano and T.Usui, Effects of Emotional Synchronization in Human-Robot KANSEI Communications, in *Proc. of the 18th IEEE International Symposium on Robot and Human Interactive Communication*

- (RO-MAN) (IEEE Press, Toyama, Japan, 2009), pp. 52-57.
14. Bursleson, W., Picard, R.W., Gender-Specific Approaches to the Development of Emotionally Intelligent Learning Companions, *IEEE Intelligent Systems, Special issue on Intelligent Educational Systems*, **22**(4)(2007)62-69.
 15. Bursleson, W., Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance, *MIT PhD Thesis*, (2006).
 16. Laurel D. Riek, Philip C. Paul, Peter Robinson, When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry, *J. on Multimodal User Interfaces*, **3**(1-2)(2010)99-108.
 17. Hyoung Rock Kim, Dong-Soo Kwon, Computational Model of Emotion Generation for Human-Robot Interaction Based on the Cognitive Appraisal Theory, *Intelligent and Robotic Systems*, **60**(2)(2010)263-283.
 18. Cornelia Wendt, Michael Popp, Berthold Faerber, Emotion induction during human-robot interaction, in *Proc. of the 4th ACM/IEEE int. conf. on Human robot interaction (HRI)* (IEEE Press, La Jolla, USA, 2009), pp. 213-214.
 19. Ching-Chih Tsai, You-Zhu Chen, Ching Wen Liao, Interactive emotion recognition using support vector machine for human-robot interaction, in *Proc. of the IEEE int. conf. on Systems, Man and Cybernetics (SMC)* (IEEE Press, Hyatt Regency Riverwalk, USA, 2009), pp. 407-412.
 20. Rachel Kirby, Jodi Forlizzi, Reid Simmons, Affective social robots, *Robotics and Autonomous Systems*, **58**(3)(2010)322-332.
 21. E. Boyle, A.H. Accomdinglyerson, A. Newlands, The effects of visibility on dialogue performance in a cooperative problem solving task, *Language Speech*, **37**(1)(1994)1-20.
 22. A. Mehrabian, Communication without words, *Psychology Today*, **2** (9)(1968)52-55.
 23. Matthew Turk, Alex Pentland, Eigenfaces for recognition, *J. of Cognitive Neuroscience*, **3**(1)(1991)71-86.
 24. Shinn-Ying Ho, Huang Hui-Ling, Facial modeling from an uncalibrated face image using a coarse-to-fine genetic algorithm, *Pattern Recognition*, **34**(5)(2001)1015-1031.
 25. Alan L. Yuille¹, Peter W.Hallinan¹ and David S. Cohen, Feature extraction from faces using deformable templates, in *Proc. of IEEE Computer Thereforec. Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE Press, San Diego, USA, 1989), pp. 104-109.
 26. Henry A. Rowley, Shumeet Baluja and Takeo Kanade, Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(1)(1988)23-38.
 27. Md. Al-Amin Bhuiyan, Vuthichai AMPORNARAMVETH, Shin-yo MUTO, Shin-yo, Haruki UENO, Face detection and facial feature extraction, in *Int. Conf. on Computer and Information Technology (ICCIT)* (ICCIT Press, Dhaka, Bangladesh, 2001), pp. 270-274.
 28. J.A. Russell, A circumplex model of affect, *Journal of Personality and Social Psychology*, **39**(6)(1980)1161-1178.
 29. H. Yamada, Models of perceptual judgment of emotion from facial expressions, *Japanese Psychological Review*, **43**(2)(2000)245-255.[in Japanese]
 30. H. Yamada, T. matsuda, C. Watari and T. Suenaga, Dimensions of visual information for categorizing facial expressions of emotion, *Japanese Psychological Research*, **35**(4)(1993)172-181.
 31. Viola P., Jones M., Rapid object detection using boosted cascade of simple features, in *Proc. of IEEE Conf. Comp. Vis. Pattern Recogn (CVPR'01)* (IEEE Press, Hawaii, USA, 2001), pp. 511-518.
 32. M. Okada, D. Nakamura, H. Kadone et al., Research on Attractor Design for the Dynamics-based Information Processing System,-On-line Design for Plastic Property and Hierarchical Design for Large-scale System-, *Journal of the Robotics Society of Japan*, **23**(5)(2005)583-593.
 33. M. Okada, K. Tatani and Y. Nakamura, Polynomial Design of the Nonlinear Dynamics for the Brain-Like Information Processing of whole Body Motion, in *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)* (IEEE Press, Washington, D.C., 2002), pp. 1410-1415.
 34. M. Okada and Y. Nakamura, Design of the Continuous Symbol Space for the Intelligent Robots using the Dynamics-based Information Processing, in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)* (IEEE Press, New Orleans, USA, 2004), pp. 3201-3206.
 35. C. Hasson, P. Gaussier, S. Boucenna, Emotions as a dynamical system: the interplay between the meta-control and communication function of emotions, *Paladyn J. of Behavioral Robotics*, **2**(3)(2011)111-125.
 36. Michael J. Lyons, Miyuki Kamachi, Jiro Gyoba, *Japanese Female Facial Expressions (JAFFE)*, Database of digital images (1997).
 37. A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, Bosphorus Database for 3D Face Analysis, in *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008)* (Roskilde University, Denmark, 2008), pp 47-56.
 38. H. Yamada, Visual information for categorizing facial expression of emotion, *Applied Cognitive Psychology*, **7**(3)(1993)257-270.



Yi Li received the B.S degree from the Southwest University, Chong Qing, China, in 2007. He is currently studying in the Department of Kansei Engineering, Shinshu University, Nagano, Japan for the M.S. degree. His current research interests include image processing, pattern recognition, and applications in human-robot communication.



Minoru Hashimoto received M.S. and Ph.D. degrees from the University of Tokyo in 1980 and 1983, respectively. From 1982 to 1988 he was a research associate of the robotics laboratory in the Department of Mechanical Engineering at the University of Electro-Communications. From 1988 to 1998 he was an associate professor in the Department of Mechanical Engineering, Kagoshima University. Since 1999 he has been a professor in the Department of Kansei Engineering, Shinshu University. His research interests are communication robots, motion assist systems, and artificial muscles.

Figures:

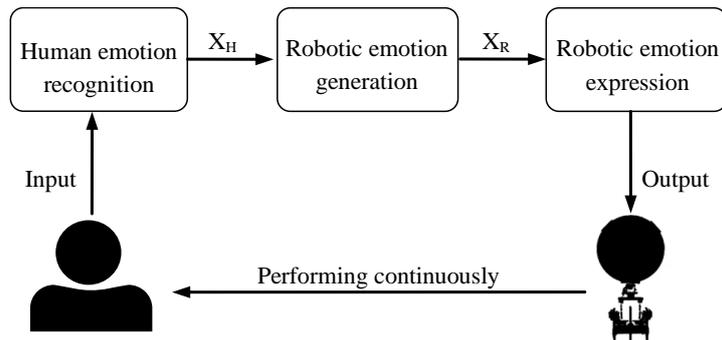


Fig. 1. The structure of the communication system.

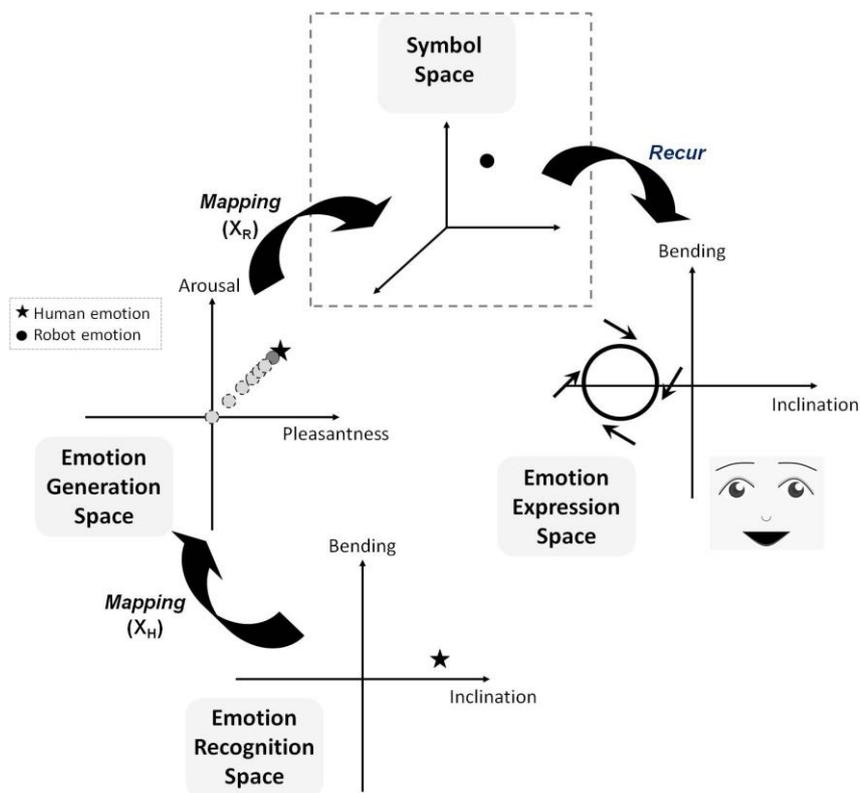


Fig. 2. Overview of the communication system.

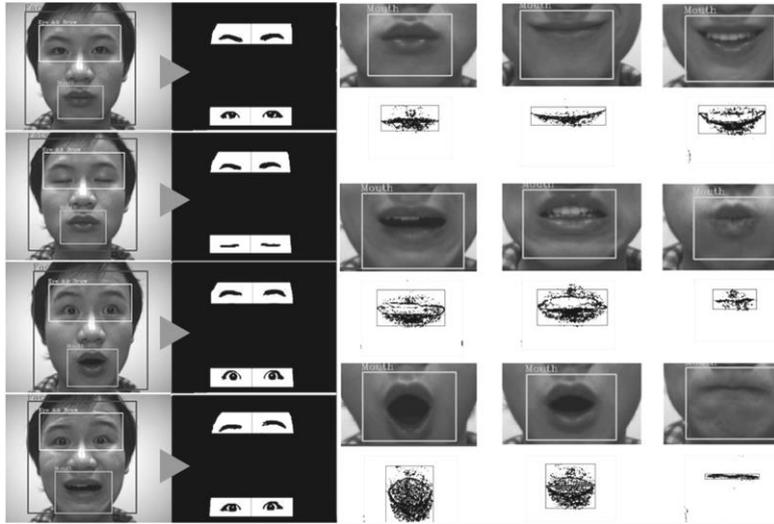


Fig. 3. Results of facial image processing.

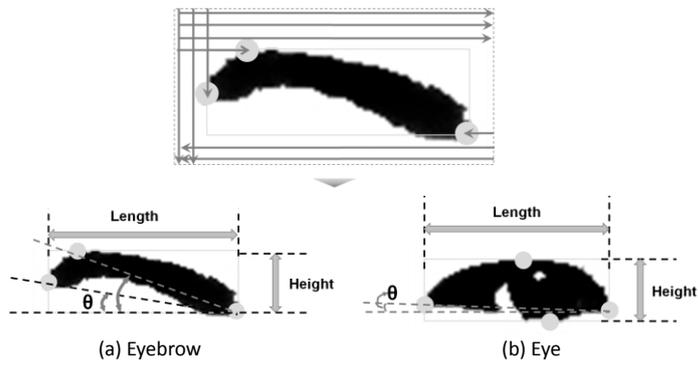


Fig. 4. Example of feature extraction of brow and eye.

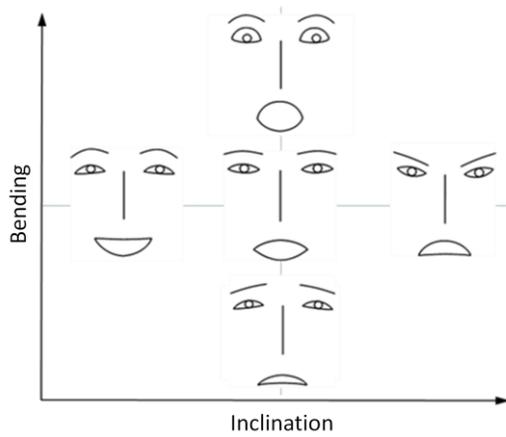


Fig. 5. Facial expressions represented by bending and inclination variables.

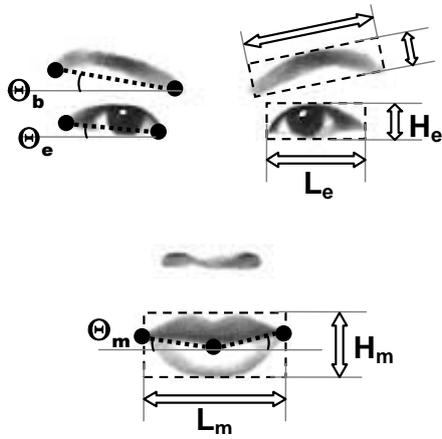


Fig. 6. Definition of bending and inclination of a facial expression.

Inclination	'+'			
	normal			
	'-'			
Bending	'+'			
	normal			
	'-'			

Fig. 7. Definition of signs for bending and inclination.

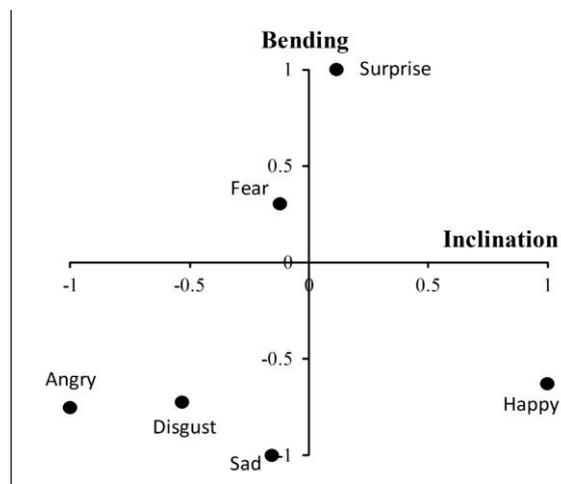


Fig. 8. 2D physical emotion recognition space.



Fig. 9. Overview of the experiment for facial expression recognition.

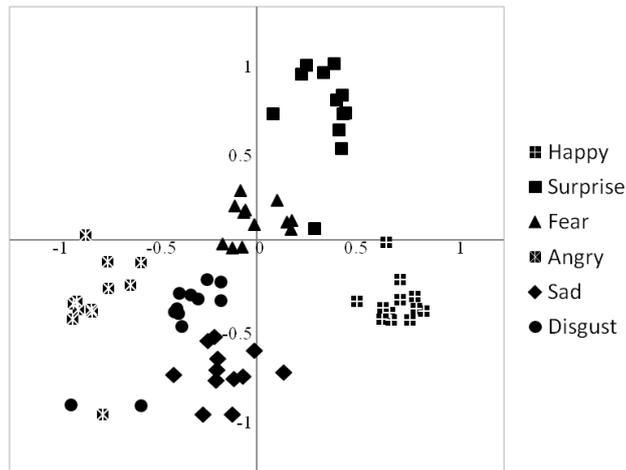


Fig. 10. Average distribution of the six basic expressions of the 12 subjects.

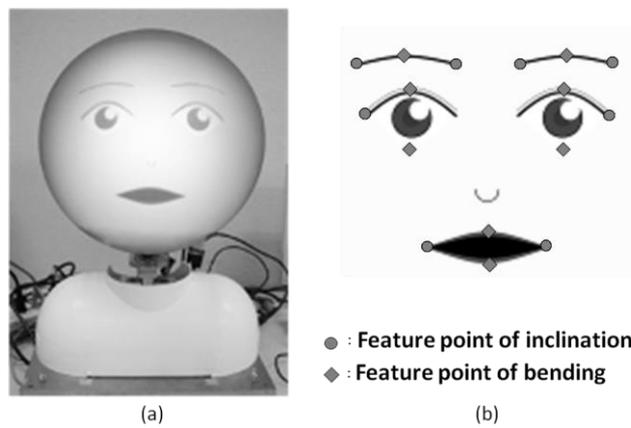


Fig. 11. Overview of the head robot KAMIN.

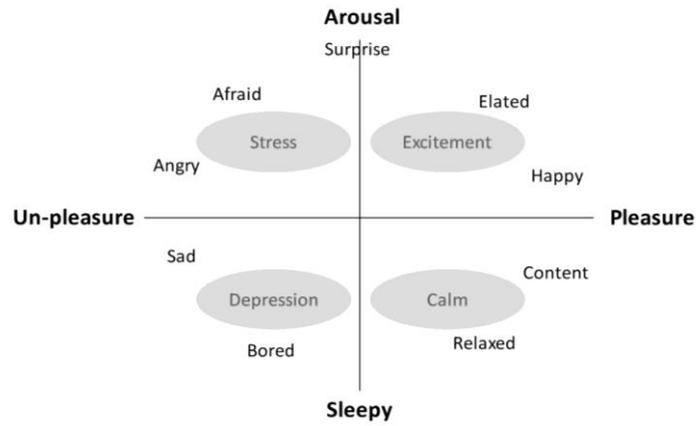


Fig. 12. Russell's circumplex model of emotion.

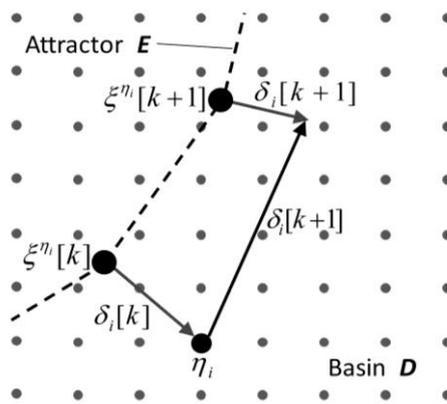


Fig. 13. Definition of the vector field.

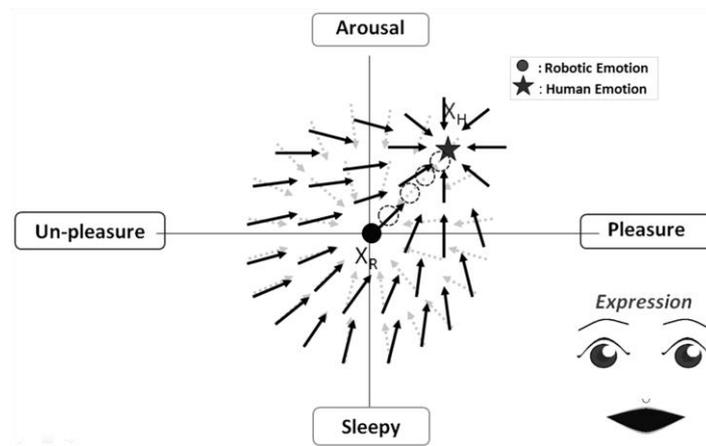


Fig. 14. Framework of the emotional synchronization.

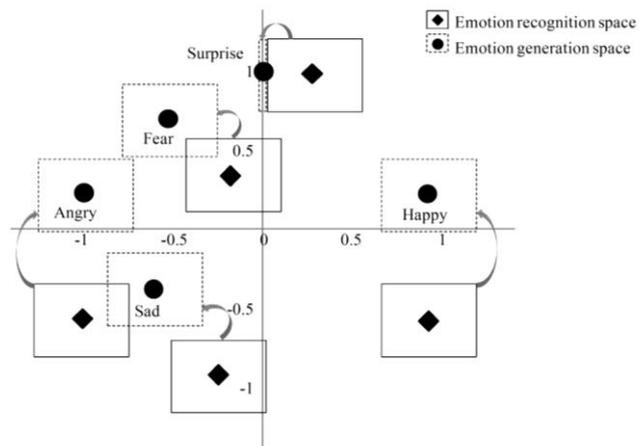


Fig. 15. Translation between emotion recognition space and emotion generation space.

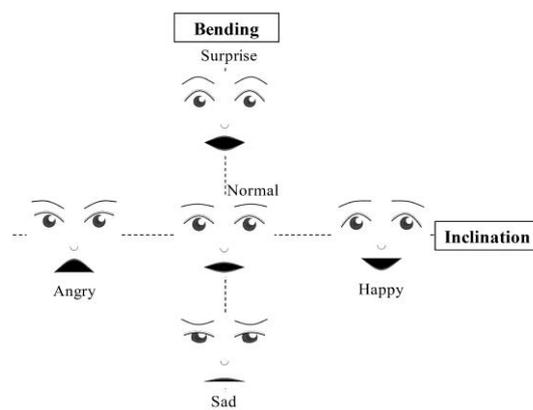


Fig. 16. An example of robotic facial expressions of basic emotions.

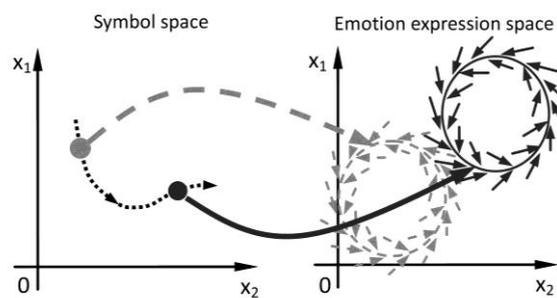


Fig. 17. The dynamic-based symbol space and emotion expression space.

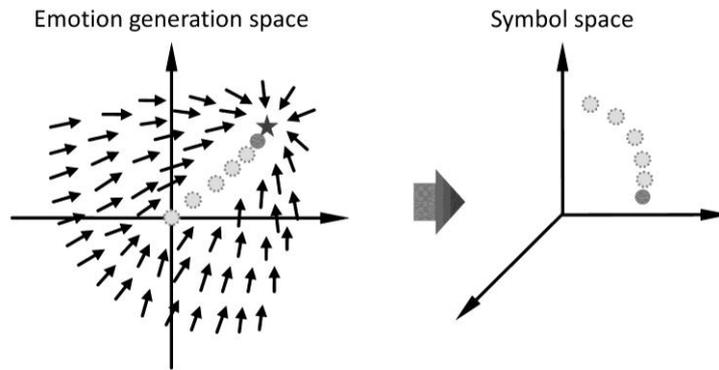


Fig. 18. Nonlinear mapping between emotion generation space and symbol space.

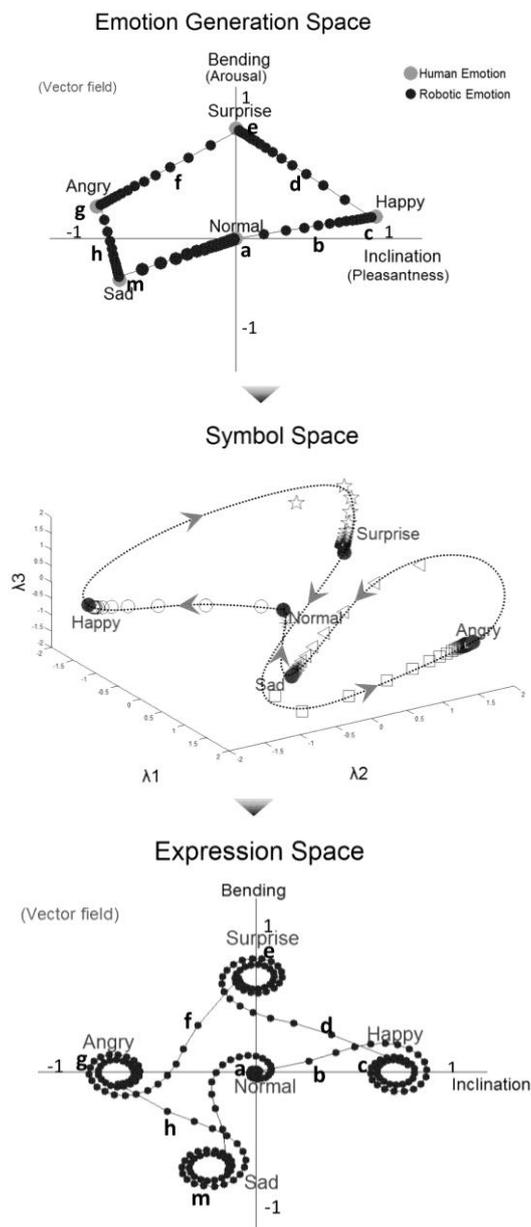


Fig. 19. Simulation of the communication system.

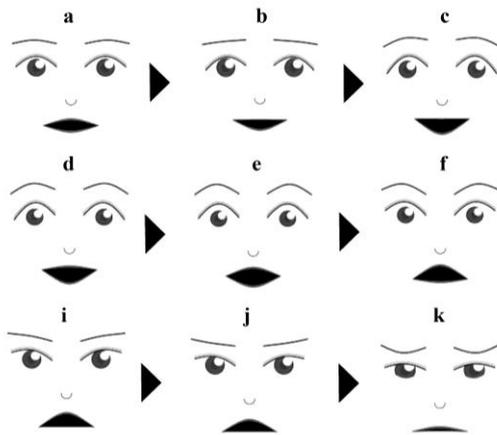


Fig. 20. The varying facial expressions of the robot.

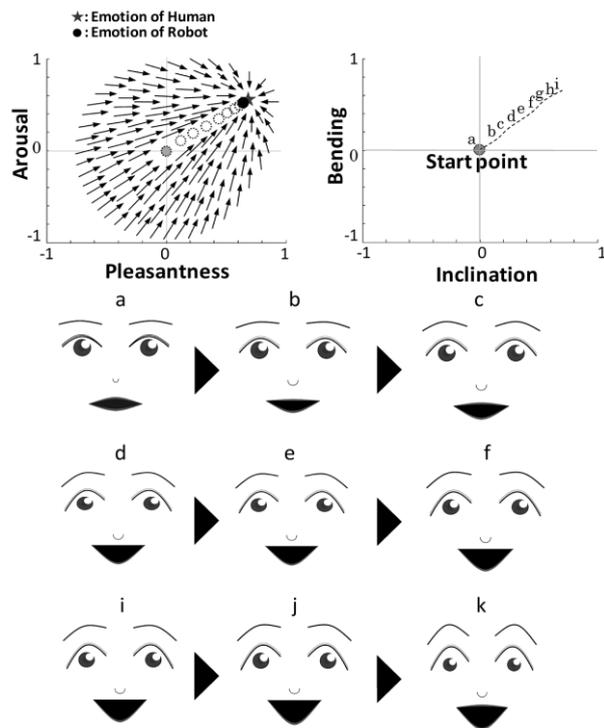


Fig. 21. An example of synchronization reaction.

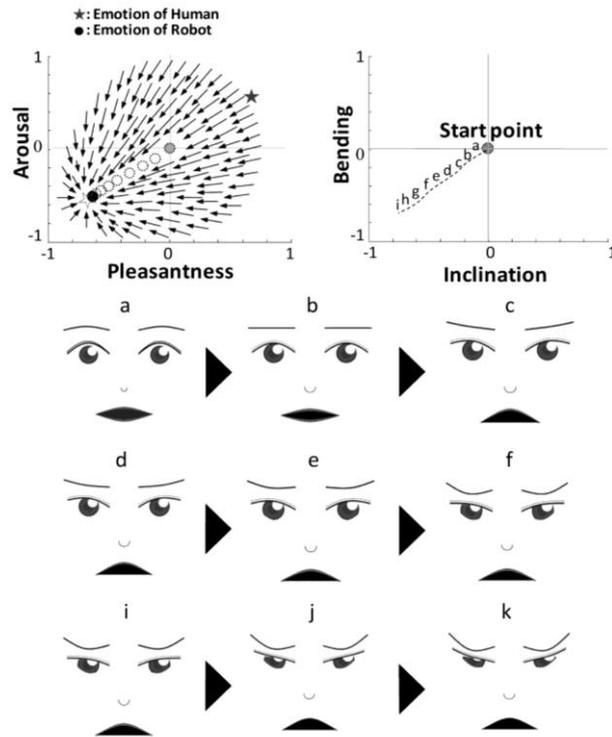


Fig. 22. An example of non-synchronization reaction.

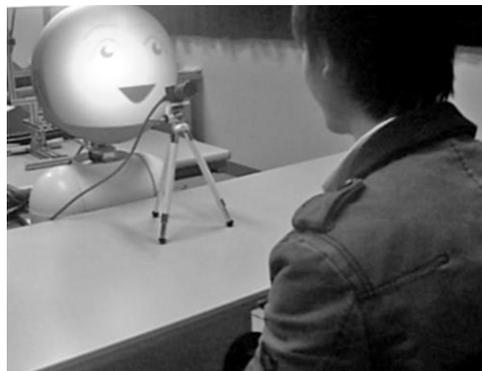


Fig. 23. Overview of the communication experiment.

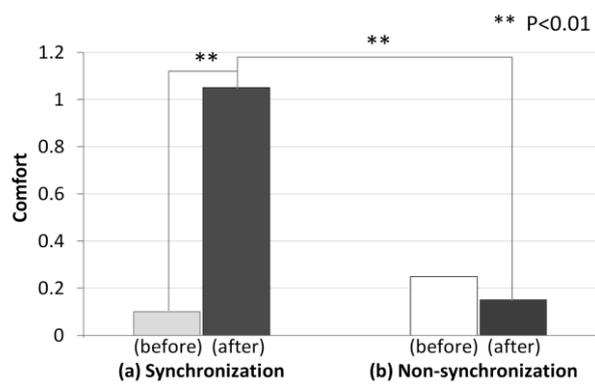


Fig. 24. Results of subjects' comfortableness and arousal after the communication experiment.

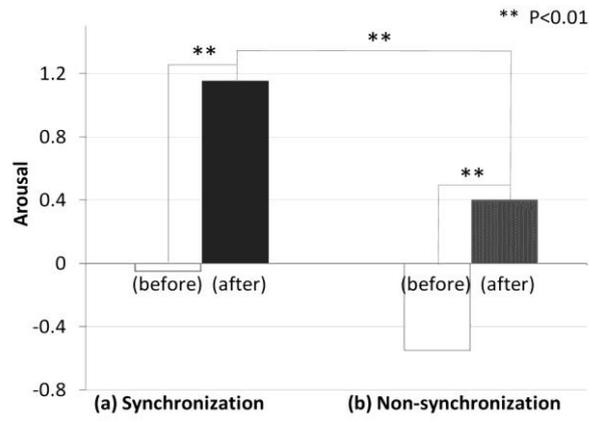


Fig. 25. Results of subjects' comfortableness and arousal after the communication experiment.

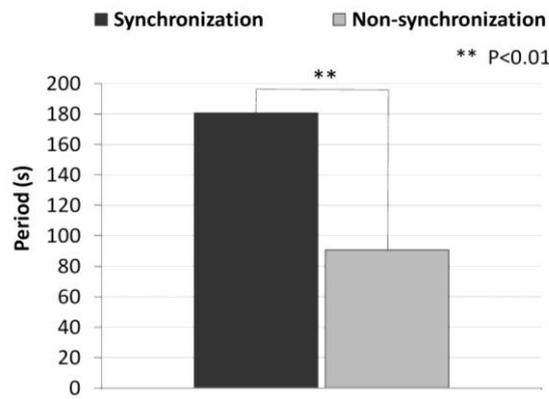


Fig. 26. Average communication time in the communication experiment.

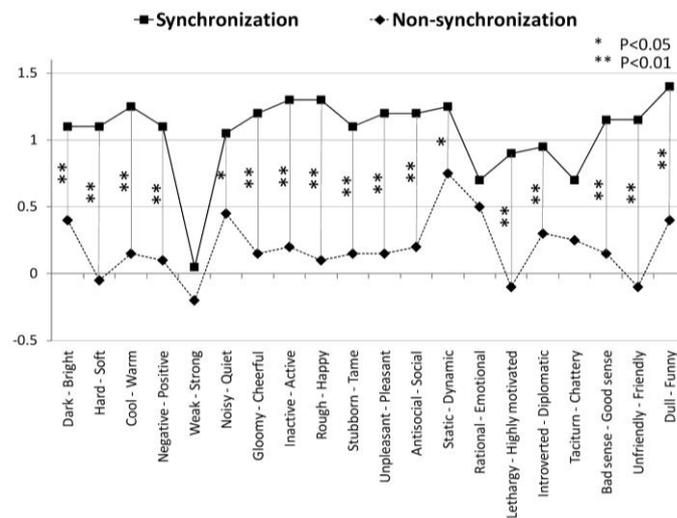


Fig. 27. Impression of KAMIN in emotional synchronized and non-synchronized communication.