INTELLIGENT AUTOMATION & SOFT COMPUTING, 2017
https://doi.org/10.1080/10798587.2017.1364919

System Integration for Cognitive Model of a Robot Partner

Jinseok Woo, János Botzheim and Naoyuki Kubota

Graduate School of System Design, Tokyo Metropolitan University, Tokyo, Japan

ABSTRACT

This paper introduces the integrated system of a smart-device-based cognitive robot partner called iPhonoid-C. Interaction with a robot partner requires many elements, including verbal communication, nonverbal communication, and embodiment as well. A robot partner should be able to understand human sentences, as well as nonverbal information such as human gestures. In the proposed system, the robot has an emotional model connecting the input information from the human with the robot's behavior. Since emotions are involved in human natural communication, and emotion has a significant impact on humans' actions, it is important to develop an emotional model for the robot partner to enhance human robot interaction. In our proposed system, human sentences and gestures influence the robot's emotional state, and then the robot will perform gestural and facial expressions and generate sentences according to its emotional state. The proposed cognitive method is validated using a real robot partner.

KEYWORDS

Intelligent robots; computational intelligence; human robot interaction; cognitive model; communication model; gesture generation; facial expression; emotional model

1. Introduction

In recent years, one of the major problems of aging society is the increasing number of elderly people who live alone. According to some statistics, by the year of 2015 the number of elderly people (65 years or older) could exceed 26.4% of the population in Tokyo (United Nations, 2012). The aging group requires daily care and human-to-human communication, but this is not a feasible option due to a lack of manpower. One of the possible solutions to overcome this social problem is to introduce human friendly robot partners to communicate and provide emotional support to the elderly people.

Many robot partners have been developed to support human life (Rane, Mhatre, & Kurup, 2014). Such human-friendly robot partners can assist humans by using voice recognition, speech, and gestural expression. Additionally, some robot partners use touch screens to display facial expressions (Romo, 2012). One of the advantages of a touch screen is that the robot does not require a hardware structure for the facial expressions, which can reduce the cost of robot partner development. In this paper, we introduce a smart device based cognitive robot partner called iPhonoid-C. The advantage of this robot is its cheap realization, as smartphones are becoming highly popular and are increasingly likely to be carried by a person (Google and conducted by Ipsos MediaCT, 2013). Since the cognitive model is implemented on a smartphone, it is therefore important to keep the computational cost as low as possible. Computational intelligence techniques can balance well between computational complexity and accuracy. The other advantage of our robot partner is that many elements of its cognitive model are based on computational intelligence. In order to reduce the computational cost for the smartphone application, we proposed a modular structure of the cognitive model in our previous papers (Botzheim, Woo et al., 2014; Woo, Botzheim, & Kubota, 2014a, b, 2015). As described in this paper, all the modules have been integrated and implemented on a smartphone. The entire integration of these modules is realized. The robot has a verbal and a nonverbal communication module. The verbal communication is used for analyzing human sentence utterance. The nonverbal communication module can recognize human gestures including human detection, motion extraction and gesture classification by applying evolution strategy, spiking neural network, and self-organizing map.

Human communication generally involves the perception of others intentions and feelings. Emotions influence actions such as incentive functions of emotion. The robot partner needs a human-like emotional mechanism, which can help to provide the meaning and value of perceptual information, and can indirectly make decisions based on the robot's internal and external state. We propose an emotional model based on the emotion-feeling-mood concept using eight feelings. The robot contains a behavior generation module to express gestural and facial information. We apply Laban Movement Analysis (LMA) and an interactive evolution strategy for expressing the robot's gestures. Next, we apply a simple fuzzy inference for facial expression generation, crucial for nonverbal communication. For verbal communication, conversation flow learning, sentence building, and a time dependent utterance system are applied for human-robot interaction.

The paper is organized as follows: Section 2 introduces the robot partner. In Section 3, the integrated system of the robot's cognitive model is presented. Section 4 discusses the verbal and nonverbal communication modules as input information to the emotional model. The emotional model is detailed in Section 5. Section 6 explains the robot's behavior generation and sentence generation techniques. Section 7 shows experimental results. Conclusions are drawn in Section 8.

CONTACT Jinseok Woo woojs@tmu.ac.jp

© 2017 TSI® Press

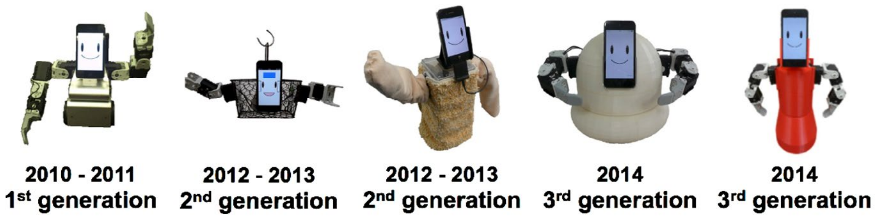


Figure 1. The History of iPhonoid.

2. Robot Partner: Iphonoid-C

As the technology develops, many smart devices are increasingly being developed for a low price and high specifications. As a result, various applications have been developed, based on the high specifications of these smart devices that are equipped with various sensors. A smart device can also be applied for developing a robot partner by utilizing the devices functions. We have developed the “iPhonoid” series as a robot partner, as shown in Figure 1. When designing the robot partner, we consider robot stability and convenience for household usage. Communication methods and degrees-of-freedom (DOF) are also factors for hardware design. Fundamentally, our system uses a smart device for realizing computational intelligence for the robot partner. By using the smart device, the traditional functions of expensive sensors are combined in one smart device.

The iPhonoid uses an iOS device, which is equipped with various sensors such as a touch sensor, microphone, two cameras, GPS, an accelerometer, a gyroscope, and a magnetometer (Apple Inc., 2016). The robot partner can be made aware of the environmental information based on information provided by the various sensors inside the robot partner.

In this paper, iPhonoid-C is introduced as a new generation of iPhonoid. The body of iPhonoid is created by a 3D-printer. If users have the 3D CAD design for printing the robot, they can create and test the robot design easily. Figure 2 illustrates the robot’s design and its size. iPhonoid consists of an iOS device, a robot body, a microcontroller, and several servomotors.

iPhonoid-C is equipped with 8 servomotors; 3 DOF on each arm and 1 DOF on its neck and waist for body movements. The actuators of the robot partner are controlled by the smartphone via Bluetooth communication (Sakata, Botzheim, & Kubota, 2013). In this paper, we apply an iPhonoid with Bluetooth 4.0 Low Energy module for compatibility with iOS (u-blox, 2017). Low energy consumption should be taken into account for the robot’s hardware design. This Bluetooth module is used for communication between iOS device and Arduino for servo control (ARDUINO, 2017). Each servomotor is controlled by the signals that are sent out from the smart device for expressing various gestures. The data structure for controlling the robots gesture is presented in Figure 3. Each value can be set between 0 and 1023; however, some values have to be excluded from this range due to limitations of the hardware structure (ROBOTIS, 2017).

In our previous work (Woo, Wada, & Kubota, 2012), we have adapted the robot partner to elderly care in a real-world environment. We realized that the robot is capable of being a good interlocutor. Thus, the robot partner requires various types of information from the real-world environment to detect the elderly people’s state. However, it is difficult to store all environmental data in real time when extracting and processing large quantities of information. Therefore, the Informationally

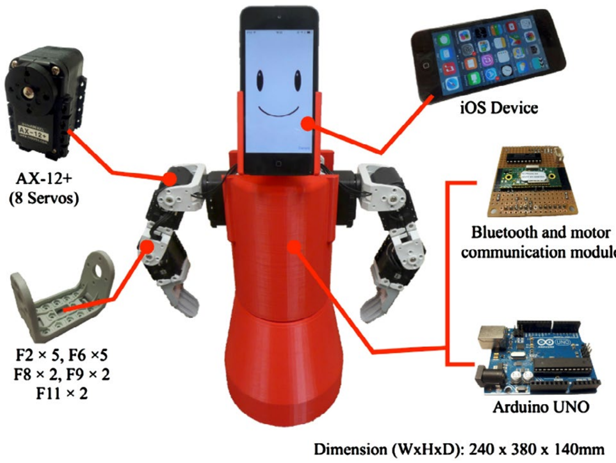


Figure 2. The Robot Partner: iPhonoid-C.

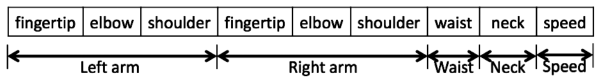


Figure 3. Data Structure for Bluetooth Communication.

Structured Space (ISS) is used as a very important tool for the robot partner to provide the elderly people care. Sensory information and the robot’s inner information are collected on the ISS server. There is a large amount of information to process on the server, which can be shared by each robot partner. The robot is able to identify the information by using the ISS server and to share information with a human. This information sharing process within the environment can realize a natural communication between a human and a robot partner (Tang, Yusuf et al., 2015).

3. Cognitive Model for the Robot Partner

Cognitive Robotics is an important technology for the robot in order to reflect the cognitive abilities of humans. In order for the robot partner to communicate with people, it is necessary to have contextual information on the environment.

Cognitive models have a long research history in psychology. The development of cognitive robotics and human cognitive systems research is interrelated with robot cognitive architecture, human-robot interaction and robot personality implementation (Breazeal et al., 2004; Burghart et al., 2005; Kurup & Lebiere, 2012).

Cognitive approaches are also used for human therapy. The generic cognitive model (GCM) has been developed for therapy of the mind and for solving psychological problems. Advances in research of the human cognitive model can have a significant effect and motivation in cognitive robotics (Beck & Haigh, 2014). Further, there has been research on integrating language to cognition (Cangelosi, Tikhanoff et al., 2007). Various artificial cognitive systems are adapted to computational agents (Vernon, Metta, & Sandini, 2007). Therefore, we can consider that cognitive architectures are important in developing the robot partner's system (Kurup & Lebiere, 2012).

The area of service robotics has introduced a robot cognitive architecture to human robot interaction. There has been a variety of discussions about robot behavior to coexist with humans (Alami et al., 2006; Dautenhahn et al., 2006). Cognitive models of our robots exhibit a personality, which is prepared by an emotion model through processing information from the outside environment by using only smart device sensory information, and we proceed to increase its practical usage.

In this paper, the robot partner system was constructed based on the cognitive model. A cognitive model depends on the inside sensors of the robot partner. A cognitive model for the robot partner is also important for realizing human-like behavior and advanced intelligence. We define a cognitive model for iPhonoid to understand and support human beings. The cognitive model of iPhonoid is depicted in Figure 4. Between perception and action, the model has 5 components:

Module 1—Nonverbal communication components e.g., gesture recognition, face detection
Module 2 - Emotion model to apply in interaction with human
Module 3 - Gesture generation by Laban theory and facial expression generation
Module 4 - Emotional state from human sentence utterance
Module 5 - Conversation modules

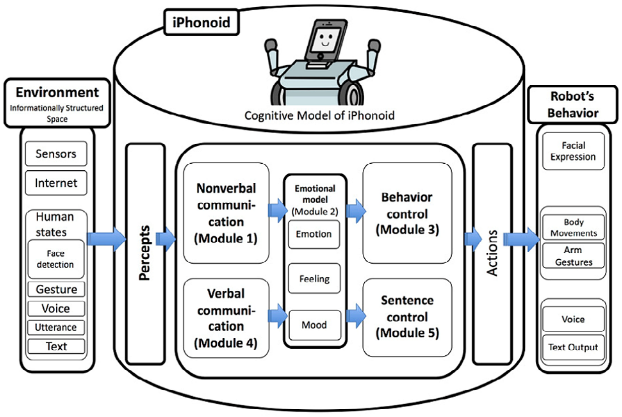


Figure 4. The Cognitive Model of iPhonoid.

The robot partner has verbal and nonverbal communication modules for estimating the robot's emotional state after perception. The robot's behavior, which includes gestural and facial expressions and the robot's sentence, will be generated based on the calculation results of the emotional model.

The verbal communication system is discussed in our previous papers (Woo & Kubota, 2013; Woo et al., 2014b). The emotional model and its relation to the nonverbal communication such as face classification and gesture recognition are explained in our other previous paper (Botzheim et al., 2014). The behavior generation module is discussed in (Woo et al., 2014a). This paper presents the integrated cognitive model based on the improved previous modules. The robot system is composed of four layers: Hardware Layer, Library Layer, Component Layer, and Application Layer (Figure 5). The service in a layer can be realized through the combination of functions based on a bottom up approach in the lower layer. By integration of the system, the service can be provided based on the locations and situations. For example, a hospital guide robot can be realized by incorporating the conversation function in the application layer. An elderly care robot can be realized based on the elderly care system function in the application layer. The modularization and integration of the system is a very important concept in iPhonoid.

4. Input Information for Cognition

4.1. Image Processing for the Robot Communication System

We use the iPhone's camera for image processing (Woo & Kubota, 2013). The robot can recognize humans for interaction based on the camera image. We use gray scale conversion, differential extraction, simple color extraction, and template matching for extracting a human shape from the camera image. First, we convert the color image to gray scale image by using the YUV model:

$$p_y(x,y,t) = 0.298912 \cdot p_R(x,y,t) + 0.586611 \cdot p_G(x,y,t) + 0.114478 \cdot p_B(x,y,t) \tag{1}$$

Where $p_R(x,y,t)$, $p_G(x,y,t)$ and $p_B(x,y,t)$ are the values of RGB at the discrete time t, respectively. The differential extraction is done by:

$$p_D(x,y) = \lambda_D \left\| p_Y(x,y,t) - p_Y(x,y,t-1) \right\| + 1 \tag{2}$$

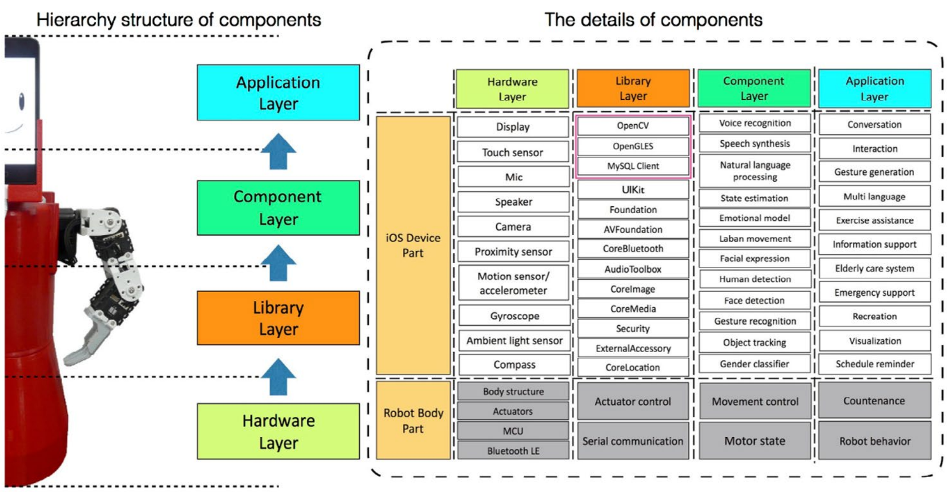


Figure 5. The Modular Architecture Layer of iPhonoid.

Where $p_D(x, y)$ is the degree of difference $(1 \le p_D(x, y) \le p_{D\_MAX})$ between two images at t and $t-1$; λ_D is a coefficient $(0 < \lambda_D < 1)$ used as a scaling factor. Furthermore, skin color is extracted as follows:

$$\begin{aligned} &if\ \lambda^G_{\min}(c) \cdot p^G(x,y,t) < p^R(x,y,t) < \lambda^G_{\max}(c) \cdot p^G(x,y,t) \\ &\text{and } \lambda^B_{\min}(c) \cdot p^B(x,y,t) < p^R(x,y,t) < \lambda^B_{\max}(c) \cdot p^B(x,y,t) \\ &\text{and } \gamma^R_{\min}(c) < p^R(x,y,t) < \gamma^R_{\max}(c) \\ &\text{then } p_{clr}(x,y) = c \end{aligned} \tag{3}$$

where c is the color ID; $\lambda^G_{\min}(c)$, $\lambda^G_{\max}(c)$, $\lambda^B_{\min}(c)$, and $\lambda^B_{\max}(c)$ are coefficients for color detection; $\gamma^R_{\min}(c)$ and $\gamma^R_{\max}(c)$ are thresholds. Thus, the robot can detect the human by using skin color extraction. For the skin color, we used the parameters in the experiments as follows: $\lambda^G_{\min}(c) = 1.2$, $\lambda^G_{\max}(c) = 2.2$, $\lambda^B_{\min}(c) = 1.2$, $\lambda^B_{\max}(c) = 2.5$, $\gamma^R_{\min}(c) = 60$.

4.2. Human and Object Detection

For the interaction between human and robot, human detection and gesture recognition are used (Woo & Kubota, 2013). In this paper, we use human gestures as input information for the robot's emotional state; because an emotional state can change, based on humans gestures. Each gesture is converted to an emotional parameter by using the robot's emotional model. Evolution strategy (ES) (Schwefel, 1981) is applied to perform human detection. We use $(\mu+1)$-ES to enhance the local hill-climbing search as a continuous model of generations, which eliminates and generates one individual in a generation. We assume that a person is moving in the image. The shape of a candidate template is generated by the $(\mu+1)$-ES. An octagonal template is used with eight searching points in the template. The i-th candidate template is represented by $g_{i,1}, g_{i,2}, \ldots, g_{i,2m+2}$ where the center of a candidate template on the image is $O_i = (g_{i,1}, g_{i,2})$; the length from the center to the j-th point is $l_j = g_{i,j+2}$; the angle is $q_j = g_{i,j+m+2}$ $(0 \le q_j \le p/4)$. The number of candidate templates (candidate solutions) is n.

The fitness value of the k-th candidate solution with color c is calculated by:

$$\begin{aligned} f_k(c) &= \sum_{(x,y)\in T_k} p_D(x,y) \cdot p_e(x,y,c) \\ p_e(x,y,c) &= \begin{cases} 1 & p_{clr}(x,y) = c \\ 0 & otherwise \end{cases} \end{aligned} \tag{4}$$

where T_k indicates a set of pixels corresponding to the k-th template. Since the result of differential extraction $(p_D(x, y))$ is used as a weighting factor, we can extract moving objects. As a result, this problem is a maximization problem.

Elitist crossover is applied, which randomly selects one individual, and generates an individual by combining genetic information between the selected individual and the best individual in order to obtain feasible solutions from the previous estimation result rapidly. If the crossover probability is satisfied, the elitist crossover is performed. Otherwise, a simple crossover is performed between two randomly selected individuals. Next, the following adaptive mutation is performed on the generated individual:

$$g_{i,j} \leftarrow g_{i,j} + \left(\beta_{1,j} \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_{2,j}\right) \cdot N(0,1) \tag{5}$$

Where f_i is the fitness value of the i-th individual, $f_{\max}$ and $f_{\min}$ are the maximum and minimum of fitness values in the population; $N(0,1)$ indicates a normal random variable; $b_{1,j}$ $(b_{1,j} > 0)$ and $b_{2,j}$ $(b_{2,j} > 0)$ are the coefficients and offset, respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population in case of maximization problems.

Human tracking is performed according to the position $(g_{m,1}, g_{m,2})$ of the m-th candidate solution. The position of the h-th human candidate in the human tracking, (X_{h1}, X_{h2}), is updated by the nearest human candidate position within the tracking range. The update is performed as follows $(k = 1, 2)$:

$$X_{h,k}(t) = (1-\alpha_H)X_{h,k}(t-1) + \alpha_H \cdot g_{m,k} \tag{6}$$

where a_H is the update rate.

4.3. Motion Extraction

In order to extract human gestures, a spiking neural network is applied (Gerstner & Kistler, 2002; Maass & Bishop, 1999). A modified simple spike response model is applied to reduce the computational cost (Botzheim et al., 2014).

The membrane potential, or internal state $h_i(t)$ of the i-th neuron at the discrete time t is given by:

$$h_i(t) = \tanh(h_i^{sys}(t) + h_i^{ref}(t) + h_i^{ext}(t)), \tag{7}$$

where $h^{syn}(t)$ includes the pulse outputs from the other neurons, $h^{ref}(t)$ is used for representing the refractoriness of the neuron, $h^{ext}(t)$ is the input to the i-th neuron from the environment. The hyperbolic tangent function is used to avoid the bursting of neuronal fires.

$$h_i^{syn}(t) = \gamma^{syn} \cdot h_i(t-1) + \sum_{j=1, j\neq i}^{N} w_{j,i} \cdot p_j(t-1), \tag{8}$$

where γ^{syn} is the temporal discount rate $(\gamma^{syn} = 0.95)$, $w_{j,i}$ is a weight from the j-th neuron to the i-th neuron, $p_j(t)$ is the pulse output of the j-th neuron at the discrete time t, and N is the number of neurons. When the internal state of the i-th neuron reaches the predefined threshold, a pulse is outputted.

The output of the i-th neuron has the value:

$$p_i(t) = \begin{cases} 1 & if\ h_i(t) \ge \theta \\ 0 & otherwise \end{cases} \tag{9}$$

where θ is a threshold for firing. When the neuron is fired, R is subtracted from $h^{ref}(t)$:

$$h_i^{ref}(t) = \begin{cases} \gamma^{ref} \cdot h_i^{ref}(t-1) - R & \text{if } p_i(t-1) = 1, \\ \gamma^{ref} \cdot h_i^{ref}(t-1) & otherwise, \end{cases} \tag{10}$$

where γ^{ref} is a discount rate and $R > 0$ $(\gamma^{ref} = 0.9)$.

The input to the i-th neuron, $h^{ext}(t)$, is calculated from the structure of the spiking neural network. The structure of neural network is directional, where eight spiking neurons are applied with a 45 degrees angle between them. Based on this encoding, the input to the i-th neuron is calculated as:

$$h_i^{ext}(t) = l(t) \cdot \exp\left(-\frac{\|\alpha_i - \alpha(t)\|^2}{\sigma^2}\right), \tag{11}$$

where α_i is the direction information of the i-th neuron, σ is the standard deviation, $\alpha(t)$ and $l(t)$ are calculated from the trajectory of the human and human hand as follows:

$$\alpha(t) = a\tan 2\left(\Delta x(t)^2 + \Delta y(t)^2\right) \tag{12}$$

$$l(t) = \tanh\left(\frac{\Delta x(t)^2 + \Delta y(t)^2}{100}\right) \cdot 0.9, \tag{13}$$

where $\Delta x(t)$ and $\Delta y(t)$ are the changes of the x and y coordinates of the moving object at time t.

4.4. Gesture Recognition

Self-organizing map (SOM) (Kohonen, 2001) is often applied for extracting a relationship among observed data, since SOM can learn the hidden topological structure from the data. Each input unit is connected to all output units in parallel via reference vectors. The input data is distributed into output units. The best matched output unit is selected according to the Euclidean distance (Botzheim et al., 2014).

The input to the SOM is given as the weighted sum of pulse outputs from neurons:

$$v = (v_1, v_2, \cdots, v_N)$$
$$v_i = \sum_{t=1}^{T} \left(\gamma^{SOM}\right)^t \cdot p_i(t), \tag{14}$$

where $p_i(t)$ is the pulse output of the i-th neuron based on Equation (9) and γ^{SOM} is a weight parameter used for distinguishing the different directions in the time. During the training phase, in every iteration a training sample (input) is used and the Euclidean distance between this input vector and the i-th reference vector of the SOM (r_i) is defined as:

$$d_i = \|v - r_i\|, \tag{15}$$

where $r_i = (r_{1,i}, r_{2,i}, \ldots, r_{N,i})$ and the number of reference vectors (output units) is M. Next, the k-th output unit minimizing the distance d_i is selected by:

$$k = \arg\min \|v - r_i\|, \tag{16}$$

Furthermore, the reference vectors are trained by:

$$r_i(t+1) = r_i(t) + \xi(t) \cdot \zeta_{k,i}(t) \cdot (v - r_i(t)), \tag{17}$$

where $\xi(t)$ is a learning rate ($0 < \xi(t) < 1$), $\zeta_{k,i}(t)$ is a neighborhood function ($0 < \zeta_{k,i}(t) < 1$) describing the relationship between the winning k-th output unit and the other output units.

The parameter setting for SOM is shown in Table 1. The learning rate and the neighborhood function decreases with time. After the training phase for any input data, the output class can be determined by selecting the nearest output unit for the given input. In this paper, two gestures are used to interact with the robot partner (Table 2). The gestures have the following meaning; hand up down is related to a happy feeling and hand left right is related to a sad feeling. The robot's emotion changes based on the human gestures by using the emotional model. The emotional model will be explained in the next section.

4.5. Verbal Communication

Previously, we proposed a learning system for sentence utterance (Woo & Kubota, 2013; Woo et al., 2014b). Essentially, we have to consider at least three important factors to realize natural communication between a human and a robot partner: (1) Conversation flow, (2) Mutual cognitive environment and (3) Words relationship. The conversation flow can be extracted from previous patterns of conversation. In this paper, we use adjective information for the emotional model. If the sentence has adjective information about emotion, the robot partner can use this information for emotional empathy with a human. When classifying the words, we use morphological analysis based on an iOS system (Apple Inc., 2011). Furthermore, we can consider the mutual cognitive environment based on the perceptual information and human behaviors. The example of adjective words is defined in Table 3 (Woo et al., 2015). This adjective parameter is defined with hands up-down gesture. In this paper, we proposed three parameter tables: (1) Adjective parameter with hand up down gesture ($i = 0$), (2) Adjective parameter with hand left right gesture ($i = 1$) and (3) Adjective parameter with other gestures ($i = 2$). The robot partner can calculate emotional states by using these rules.

Table 1. Parameter Settings for SOM.

iter.	N	M	init. range	γ^{SOM}	ξ	τ	ζ_1	ζ_2	ζ_3
2000	8	10	0.01	0.98	0.3	1000	0.9	0.7	0.5

5. Emotional Model for Interaction

The relationship between human emotion and communication has long been discussed from many viewpoints (Bartneck & Reichenbach, 2005). We apply the concepts of emotion, feelings, and mood based on a time scale, assuming that emotions change temporally based on the human sentence information, on the internal state, and on the external environment (Botzheim et al., 2014; Yorita, Botzheim, & Kubota, 2013). Emotion is considered as an intense short-term mental state based on perceptual information and used as intermediate input from the perceptual system to the emotional model. In this paper, we used adjective information as the emotion parameter. Each feeling is updated as the summation of emotion parameters.

The (j, k) emotional input $u^E_{j,k}(t)$ is generated based on the adjective word and gesture information as follows:

$$u^E_{j,k}(t) = \begin{cases} 1 & \text{if } (S_{adjective} = S_j) \wedge (G_{gesture} = G_k) \\ 0 & otherwise \end{cases} \tag{18}$$

where $S_{adjective}$ is the adjective in the human's sentence and S_j is the j-th adjective related to the (j, k) emotion and $G_{gesture}$ is the human gesture and G_k is the k-th gesture. The adjective words are classified into eight groups, one group for each feeling, and six adjective words are included in each adjective set. The gesture information also has feelings based on the robot personality. Each adjective word and gesture has an ID number used for calculation of the emotion parameter.

In Yorita et al. (2013) five different feeling models have been proposed. In this paper, we use one of these models, where the state of the i-th feeling $u^F_i(t)$ is updated by the emotional input from the viewpoint of bottom-up construction and the top-down constraints from mood values are also considered as displayed in Figure 6:

$$u^F_i(t) = \tanh(\kappa u^F_i(t-1) + (1-\kappa)[E_i + F_i]), \tag{19}$$

Table 2. The Definition of Gestures.

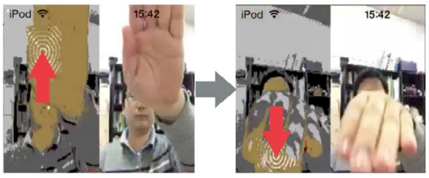
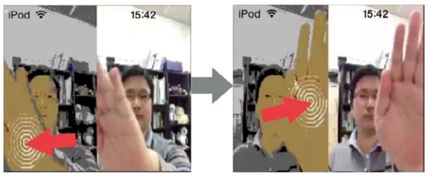
Gesture definition	Gesture image
Hand up and down gesture to robot	
Hand left and right gesture to robot	

Table 3. Coefficients between Emotions and Feelings ($w^E_{i,j,k}$).

Words	Neutral	Happy	Surprised	Angry	Disgusted	Sad	Frightened	Fearful	Thrilling
Neutral	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Happy	0.0	1.0	0.8	0.0	0.0	0.1	0.1	0.0	0.7
Nice	0.0	0.9	0.7	0.0	0.0	0.0	0.1	0.0	0.6
Glad	0.0	0.9	0.6	0.0	0.0	0.0	0.1	0.0	0.5
…	…	…	…	…	…	…	…	…	…
Breathtaking	0.0	0.2	0.1	0.0	0.0	0.3	0.0	0.2	0.4

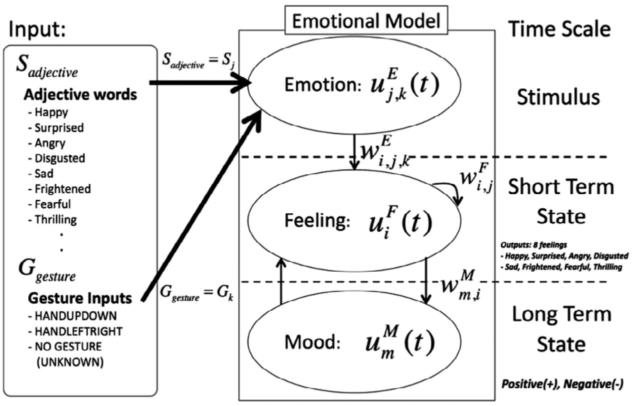


Figure 6. The Structure of the Proposed Emotional Model.

Where

$$E_i = \sum_{j=0,k=0}^{N^E} w^E_{i,j,k} \cdot u^E_{j,k}(t-1)$$
$$F_i = \sum_{j=0,j\neq i}^{N^F} w^F_{i,j} \cdot u^F_j(t-1) \quad (20)$$
$$\kappa = \frac{\gamma^F}{1 + u^M_1(t-1) - u^M_2(t-1)}$$

Where γ^F is the temporal discount rate of feelings ($0 < \gamma^F < 1$, $\gamma^F = 0.97$), N^E is the number of emotional inputs (number of adjectives: 48, number of gestures: 3), $w^E_{i,j,k}$ is the weight between the (j, k) emotion and i-th feeling ($0 \le w^E_{i,j,k} \le 1$, Table 3), N^F is the number of feelings, 8, $w^F_{i,j}$ is the stimulation or suppression coefficient from the j-th feeling to the i-th feeling ($0 \le w^F_{i,j} \le 1$), and $u^M_m(t)$ is the value of the m-th mood. Here we use positive mood ($m = 1$) and negative mood ($m = 2$) (Woo et al., 2015). The hyperbolic tangent is used to regulate the values of feelings.

Mood is defined as the long-term state updated by a change in feelings, and governs changes in feelings. Feeling is defined as a short-term state updated by a change in emotion. The state of the m-th mood is updated by the sum of feelings:

$$u^M_m(t) = \tanh\left[\gamma^M u^M_m(t-1) + (1-\gamma^M)\sum_{i=1}^{N^F} w^M_{m,i} u^F_i(t)\right], \quad (21)$$

Where γ^M is the discount rate ($\gamma^M = 0.9$) and $w^M_{m,i}$ is the stimulation or suppression coefficient from the i-th feeling to the m-th mood ($0 \le w^M_{m,i} \le 1$). The structure of the model is shown in Figure 6. In this figure, we can see how the feeling and mood influence each other and the emotion can be considered as an input impulse to the feeling. The robot has nine feelings. One is neutral for normal state others are feeling for realizing the robot's emotional model.

The relationship and arrangement between feelings are illustrated in Figure 12. The normal state of the robot is Neutral. Rules pertaining to facial expression changes are shown in Section 6.4. The robot's eight feelings are as follows: Happy, Surprised, Angry, Disgusted, Sad, Frightened, Fearful, Thrilling. The feeling states have their complementary states shown as follows:

- Happy and Sad
- Angry and Fearful
- Surprised and Frightened
- Disgusted and Thrilling

This feeling structure is inspired by Plutchik's wheel of emotions model (Plutchik, 2001). Plutchik's wheel of emotions model considers eight primary emotions such as; joy, trust, fear, surprise, sadness, disgust, anger, anticipation.

6. Output Information for Interaction

Previously, human emotion was used to adapt in the industrial area. Kansei Engineering (KE) is one of the fields to develop the human friendly system (Nagamachi, 1995). This field could support new designs of goods to enhance convenience of usage based on questionnaire information. Kansei Engineering functions by using human emotion, whereas the difference with our system is that we used emotional factors to generate robot emotion.

In human society, human decision has a relation with emotional state, even if the result is not the best solution. Consequently, an emotion model is considered in order to make a robot a human friendly system. This section presents how the utterance, facial and gestural expressions are shown in accordance with the emotional and mood difference.

6.1. Utterance System for Robot Partner

The sentence generation in the cognitive model of iPhonoid (Figure 4) consists of three subsystems: The conversation flow utterance system, the sentence building utterance system, and the time dependent utterance system. The flowchart of the conversation system is illustrated in Figure 7.

(1) Conversation Flow Utterance System (CFUS): The conversation flow can be extracted from previous patterns of the conversation (Woo & Kubota, 2013). We improved the sentence selection method by using word relationships for verbs and adjectives (Woo et al., 2014b). The conversation learning is performed by imitating the human's utterance patterns. In the beginning of the conversation learning, the robot partner has no conversation utterance. When the person speaks a sentence, the robot partner memorizes the sentence as the l-th sentence in the utterance sentence list ($l = 1, 2, \ldots$). Next, if the recognized human utterance sentence is matched with the database of utterance sentence lists, the selection strength and weight parameter are updated for the next conversation. Additionally, each robot sentence has an emotional state from the human sentence based on the emotional model (Woo et al., 2015). Therefore, the robot partner is able to generate speech based on its mood value. The robot is able to talk in a human-like manner in this way by using many sentences from the database of utterance sentences.

(2) Sentence Building Utterance System (SBUS): In our previous research (Woo & Kubota, 2013) we realized the need for the robot to generate a sentence depending on the situation. We proposed sentence building rules based on the robot's mood (Woo et al., 2015). For example, when the human says, "What is your name?" the robot has two rules based on the mood state. When the robot's mood is good, the robot says, "My name is iPhonoid-C". However, when the robot's mood is bad, the robot says "What is my name?" based on the sentence building rule.

(3) Time Dependent Utterance System (TDUS): In the case of time dependent utterance system, the robot uses contents from the database system. We prepare two-time dependent utterance contents based on the mood state of the robot partner. The two contents have different moods but a similar story. For example, if the robot's mood is good then the robot will say, "Bathing is important for health". However, when the robot's mood is bad, the robot will say, "Bathing is important to reduce stress".

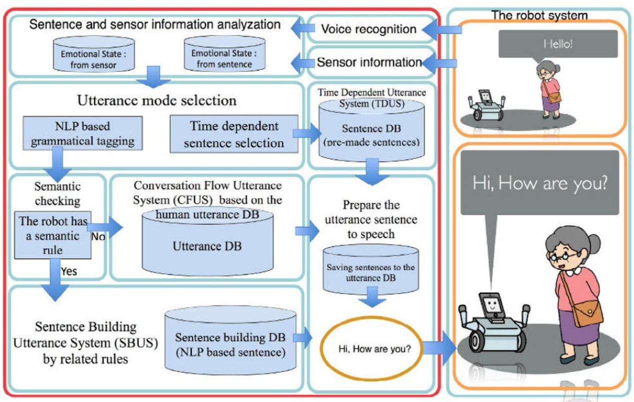


Figure 7. The Flowchart of the Conversation System.

6.2. Gesture Generation Based on Laban Theory

Laban Movement Analysis (LMA) is a theory to describe and interpret various human movements (Laban, 1980; Laban & Lawrence, 1947). LMA investigates the processes underlying human movements. LMA describes four movement components; Body, Effort, Shape, and Space (Lourens, Van Berkel, & Barakova, 2010). Efforts are those processes in the human movement that express subjective inner intention. These efforts have the following dimensions, where each one has two polarities: Space, Weight, Time, and Flow - as explained in Table 4. The combination of three factors is a drive. Action drive is an important combination, which considers the Space, Weight, and Time factors. In our previous paper, LMA was applied

Table 4. Efforts and Action Drive in Laban Movement Analysis.

Effort Factor	Effort element (Indulging Polarity)	Effort element (Fighting Polarity)	Expression method in the robot
Space	Flexible	Direct	Angle of each joint
Weight	Light	Strong	Speed of motors
Time	Sustained	Quick	Interval of timing
Flow	Free	Bound	(Not applicable)

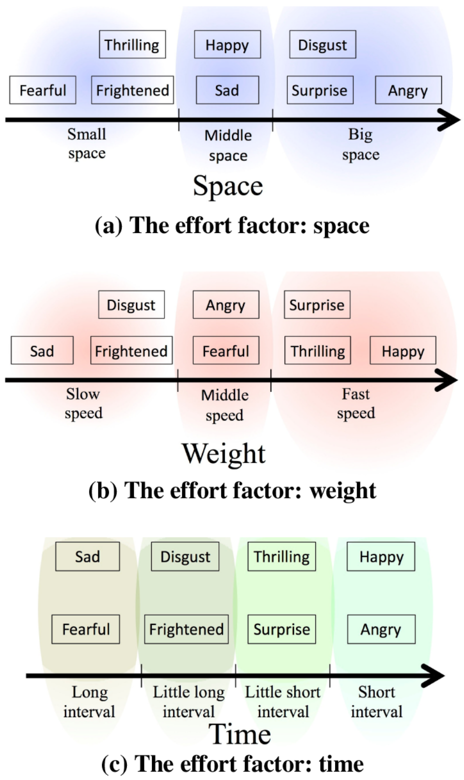
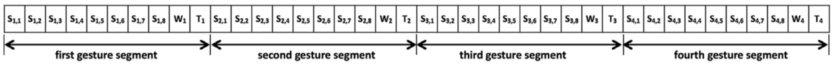


Figure 8. The Relationship between Emotions and Effort Factors.

Table 5. Efforts in Laban Movement Analysis.

	Neutral	Happy	Surprised	Angry	Disgusted	Sad	Frightened	Fearful	Thrilling
Weight	Normal	Fast	Fast	Middle	Slow	Slow	Slow	Middle	Fast
	0.0~0.5	0.5~1.0	0.5~1.0	0.3~0.7	0.0~0.5	0.0~0.5	0.0~0.5	0.3~0.7	0.5~1.0
Time	Normal	Short interval	Little short interval	Short interval	Little long interval	Long interval	Little long interval	Long interval	Little short interval
	0.0~0.5	0.0~0.25	0.25~0.5	0.0~0.25	0.5~0.75	0.75~1.0	0.5~0.75	0.75~1.0	0.25~0.5
Space	Normal	Middle	Big	Big	Big	Middle	Small	Small	Small
	0.0~0.5	0.3~0.7	0.5~1.0	0.5~1.0	0.5~1.0	0.3~0.7	0.0~0.5	0.0~0.5	0.0~0.5



	Left fingertip	Left elbow	Left shoulder	Right fingertip	Right elbow	Right shoulder	Waist	Neck	Motor Speed	Gesture interval
1st segment	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	$S_{1,4}$	$S_{1,5}$	$S_{1,6}$	$S_{1,7}$	$S_{1,8}$	W_1	T_1
2nd segment	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$	$S_{2,4}$	$S_{2,5}$	$S_{2,6}$	$S_{2,7}$	$S_{2,8}$	W_2	T_2
3rd segment	$S_{3,1}$	$S_{3,2}$	$S_{3,3}$	$S_{3,4}$	$S_{3,5}$	$S_{3,6}$	$S_{3,7}$	$S_{3,8}$	W_3	T_3
4th segment	$S_{4,1}$	$S_{4,2}$	$S_{4,3}$	$S_{4,4}$	$S_{4,5}$	$S_{4,6}$	$S_{4,7}$	$S_{4,8}$	W_4	T_4

Figure 9. Chromosome Structure of Motors for each Feeling in the Robot's Gesture.

to analyze the gesture produced by a simulator robot, and to estimate the real robot's gesture (Nishimura, Kubota, & Woo, 2012; Woo et al., 2014a).

In this paper, we used LMA for gesture generation. The robot's gesture and body movement consist of four gesture segments. Each gesture segment has the structure presented in Figure 3. Between the motion segments, a delay is applied. Table 4 presents how the Action drive concept is used in the gesture expression. The Space property is related to the joint angles encoded in the first eight values of each four-gesture segments. The Weight factor reflects the motor speed expressing the strength of the gesture encoded in the 9-th value of each gesture segment. The Time factor will describe the continuity of the gesture in our proposed model. This value is encoded between the gesture segments and its possible value is between 1 and 3 s.

We consider emotional state when making robot gestures, because there is a close relationship between emotion and body movements (Morita, Nagai, & Moritsu, 2013). The robot can generate its movement based on its feeling using the effort factors in Action drive. Figure 8 illustrates the relationship between the eight feelings and the effort factors. The robot's feelings depend on its perception, which is related to verbal or nonverbal communication as explained in the cognitive model of iPhonoid (Figure 4).

This relationship is further detailed in Table 5 by using Figure 8. Although these relations are subjective, they reflect well to our everyday experiences. For example, the happy feeling has much faster and more continuous movement than the fearful feeling, and the gesture related to the angry feeling needs bigger space than the other states do. The relationships are also quantified in Table 5. The intervals in Table 5 are used to generate the initial gesture for the given feeling. Each parameter is shown as a range of values. Generally, humans have different gesture shapes to express specific emotion (Castellano, Villalba, & Camurri, 2007). Therefore, we tried to select basic parameter of LMA based on effort factors (Figure 8, Table 5).

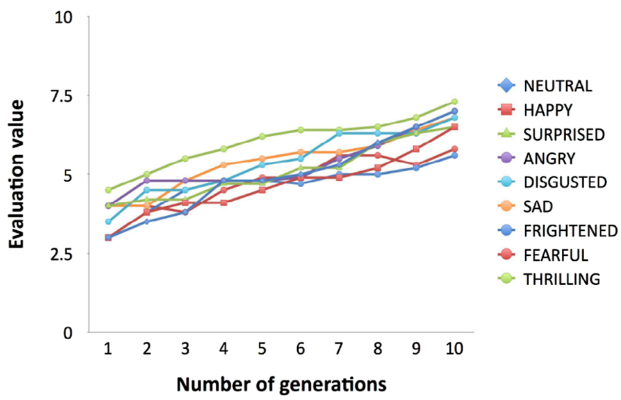


Figure 10. Evaluation Results by IEC.

6.3. Gesture Optimization by Interactive Evolution Strategy

In terms of Kansei Engineering, robot gesture evaluation is achieved through relative evaluation by a human on the robot gesture. Here, we select robot gesture of LMA by using the IES evaluation on one of the best robot gestures by using human evaluation. The initial gesture expression obtained by Laban theory is optimized by Interactive Evolution Strategy (IES). In Interactive Evolutionary Computation (IEC), human evaluation is applied, because the fitness function of the candidate solutions is not known (Sims, 1992; Takagi, 2001). We apply this approach for evaluating the robot's gesture. The person can give a numerical score to indicate the quality of the gesture. Evolution strategy is applied as the evolutionary computation technique (Schwefel, 1981). A simple (1+1)-ES approach is applied, where one parent competes with its mutated offspring. The fittest individual from the parent and the offspring will survive to the next generation. The mutation in the offspring is performed as follows:

$$x_{t+1} = x_t + N(0, \sigma), \tag{22}$$

Table 6. Motor Parameters for Each Feeling.

	Neutral	Happy	Surprised	Angry	Disgusted	Sad	Frightened	Fearful	Thrilling
$S_{1,1}$	0.047974	0.491513	0.535329	0.592346	0.737727	0.585659	0.444423	0.090898	0.0
$S_{1,2}$	0.186908	0.512545	0.516291	0.821805	0.727692	0.631163	0.44184	0.270745	0.337895
…	…	…	…	…	…	…	…	…	…
T_4	0.162751	0.200234	0.326966	0.15791	0.598809	0.760632	0.745242	0.832804	0.349876

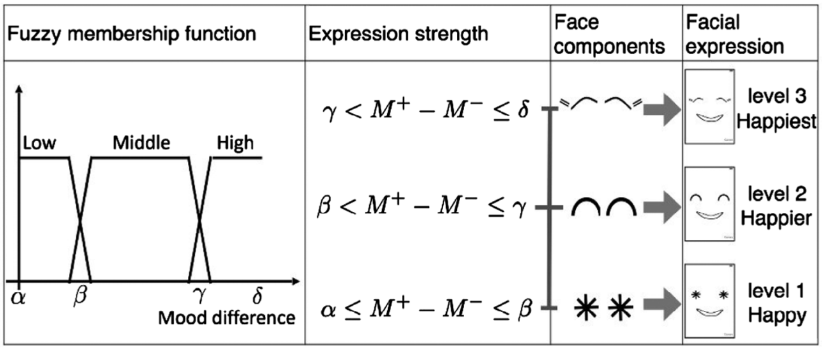


Figure 11. A Simple Fuzzy Model for Facial Expression.

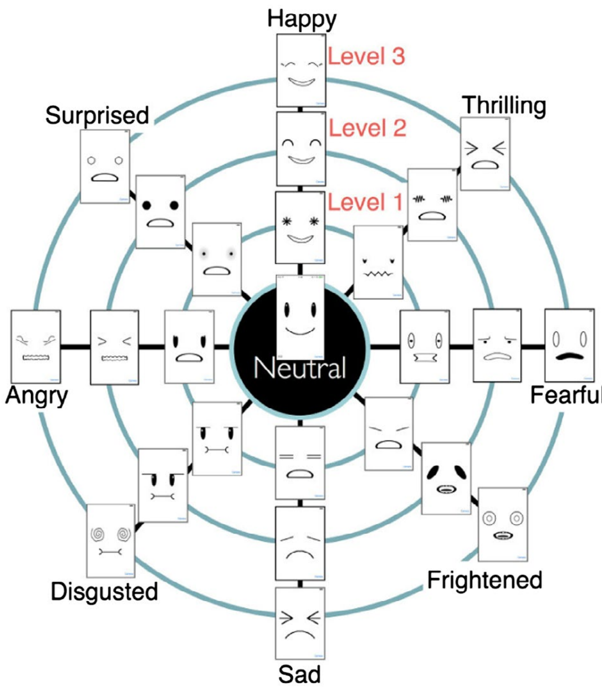


Figure 12. Facial Expression for each Feeling State.

Where x_t is the parent, x_{t+1} is the offspring, and $N(0,\sigma)$ is a normal random number with 0 mean and σ variance. The parent and the offspring have the chromosome structure shown in Figure 9, so they are 40 dimensional vectors with real components between 0 and 1. There is only one σ variance parameter for all 40 components. We follow the 1:5 rule to change the σ parameter adaptively. If more than one of five mutations are successful (i.e. the offspring is fitter than the parent) then σ is multiplied by a parameter a ($a > 1$). If the rate of successful mutations is less than 20% then σ is divided by a. Ten generations are used in the IES with σ = 0.1 and α = 1.22 parameters. The fitness evolution for the eight feelings is shown in Figure 10. As a result, the robot can perform the gesture for each feeling based on the motor parameters in Table 6. The parameters are generated based on Table 5.

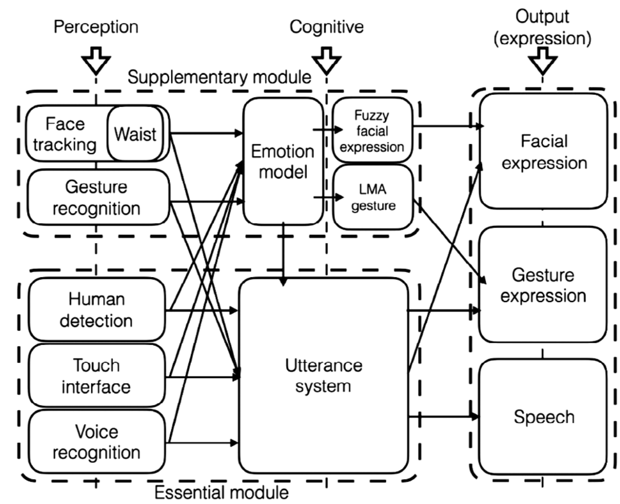


Figure 13. Functional Block Diagram of Cognitive System.

6.4. Facial Expression Generation

Human facial expression is good material of emotional contagion (Frith, 2009). Therefore, the robot's facial expression is an important factor to make a human friendly robot partner. In the facial expression generation, we select the face from a given set depending on the feeling of the robot. The mood value of the robot is also considered when selecting the face for the robot's feeling from a pre-programmed face set. A simple fuzzy approach is used based on the difference of the positive mood value and negative mood value of the robot. An example fuzzy rule base for the happy feeling is given as follows and illustrated in Figure 11. The result of the fuzzy inference is used to select from the pre-programmed happy, happier, and happiest faces. M is the first letter of mood. The parameters of the membership functions can be different for each feeling and they are parameters of the iPhonoid's cognitive model:

IF mood difference is Low, THEN Happy facial expression.

IF mood difference is Middle, THEN Happier facial expression.

IF mood difference is High, THEN Happiest facial expression.

Table 7. The Combination of Modules (○: Must, Δ: Considerable, ×: No need).

Index	Modules	Case 1(7.1)	Case 2(7.2)	Case 3(7.3)
Essential Perception	Human detection	○	○	○
	Voice recognition	○	○	○
	Touch interface	○	○	○
Supplementary perception	Face tracking	Δ	×	○
	Gesture recognition	○	Δ	Δ
Essential cognitive	Utterance system	○	○	○
Supplementary cognitive	Emotion model	○	×	○
	Facial expression	○	×	○
	LMA Gesture	○	×	Δ
Essential output	Facial expression	○	○	○
	Gesture expression	○	○	○
	Speech	○	○	○
Supplementary output	Waist	×	×	○

Table 8. Interaction Results.

State	Human Input	Robot Output	Scene	Utterance system	Emotional state	Robot gesture	Relevant Module
–	The robot stays in the room.		Figure 14-(I)	–	–	–	–
Start	(Human detection: skin color)	Hello.	Figure 14-(II)	–	–	–	• Human detection
HS1	Hi!(Human gesture: hand up down)	How are you doing?	Figure 14-(III), (IV)	CFUS	Happy	LMA based gesture	• Gesture recognition • Voice recognition • Emotion model • Utterance system
HS2	I'm fine, thank you.	I'm good, thank you.	Figure 14-(IV)	CFUS	Neutral	LMA based gesture	• Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
HS3	I'm sleepy.	You are welcome.	Figure 14-(IV)	CFUS	Neutral	LMA based gesture	• Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
HS4	Really?	Let's hang out today!	Figure 14-(IV)	CFUS	Neutral	LMA based gesture	• Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
HS5	Okay, I will go out.	Yes, please.	Figure 14-(IV)	CFUS	Neutral	LMA based gesture	• Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
HS6	See you later.	Yes.	Figure 14-(IV),(V)	SBUS	Neutral	Nodding head	• Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
HS7	Bye-bye.	Bye-bye.	Figure 14-(VI)	CFUS	Sad	LMA based gesture	• Gesture recognition • Voice recognition • Emotion model • Utterance system • Facial expression • Gesture generation
End	The human leaves the room. (Human detection: None)		–	–	Sad	–	–

Table 9. Interaction Results.

State	Human Input	Robot Output	Utterance system
–	Human and robot stay in the room.		–
HS1	(Human detection:Yes)	Do you want to check today's schedule?	TDUS
HS2	Yes.	We have a meeting after dinner at room number 5.	TDUS
HS3	Thank you.	I like it so much.	CFUS
End	The human leaves the room.(Human face detection: None)		–

Table 10. Interaction Results.

State	Human Input	Robot Output	Scene
–	Human and robot stay around the table.		Figure 19
Start	(Human face detection: Yes)	Hi, Would you like a cup of tea?	Figure 19-(I)
HS1	Human 1: "No."	Okay. (waist movement)	Figure 19-(I)
HS2	(Human face detection: Yes)	Hi, Would you like a cup of tea?	Figure 19-(II)
HS3	Human 1: "Yes."	Okay. (waist movement)	Figure 19-(II)
End	The human waits around the table.		–

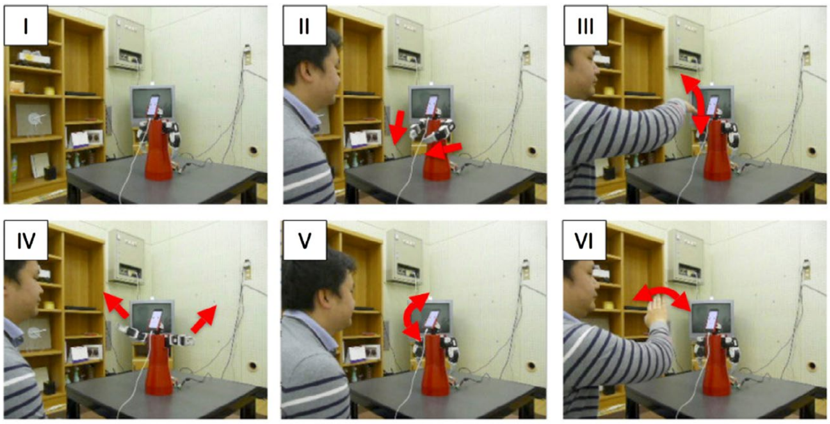


Figure 14. Experiment of Human Robot Interaction.

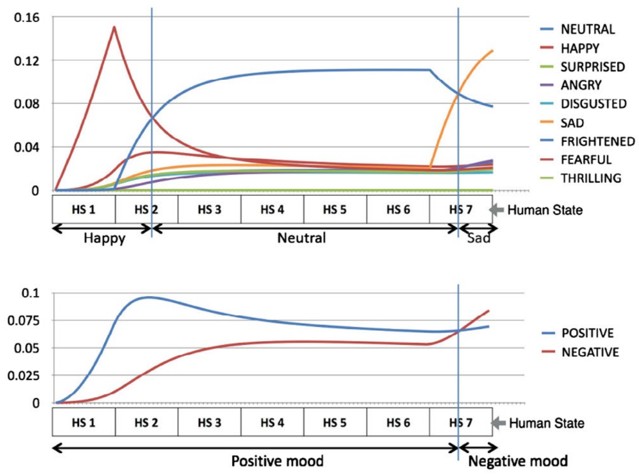


Figure 15. The Result of Robot Emotional State: The Transition of Feeling and Mood.

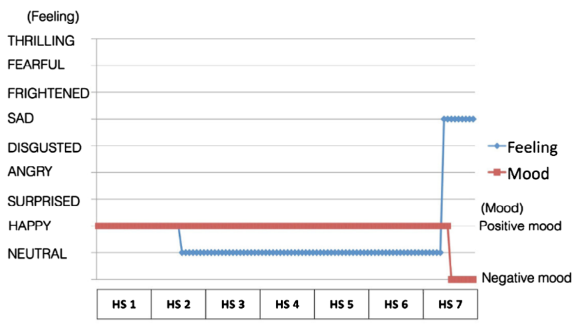


Figure 16. The Result of Robot Emotional State: Classification of Feeling and Mood.

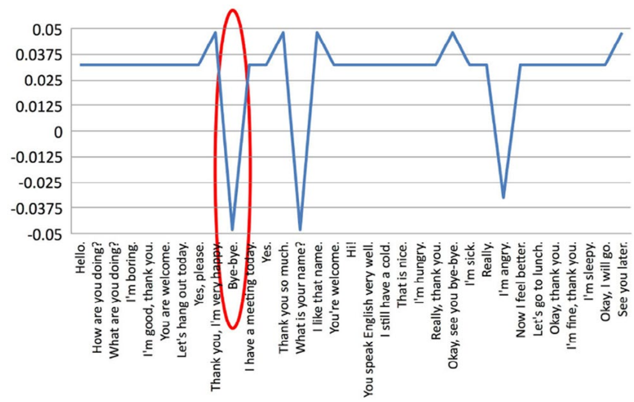


Figure 17. The Result of Robot Utterance: The Sentence Selection in the Negative Mood.

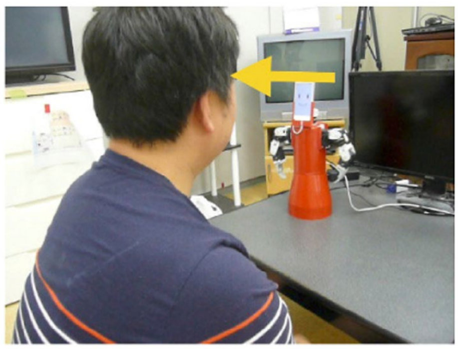


Figure 18. Experiment of Information Support.

After the calculation of mood difference between positive (M^+) and negative (M^-) parameter, it is used to select the face parts for adjusting the expression strength level as shown in Figure 11. Therefore, the robot has three different facial expressions of each feeling based on the expression strength (Figure 12).

7. Experimental Results

This section shows experimental results by using the proposed system for system integration based on modular architecture (Figure 5). We show three examples of the experiment by module as changing the system module, namely information service at home, concierge service at the event venue and staff supporting service in restaurant. The details of modules needed for the robot system configuration is shown in Figure 13. The basic module of robot consists of human detection, touch interface and voice recognition and supplementary module consists of face tracking, gesture recognition, and emotion model. The organization of modules to each experiment is shown in Table 7. We show three experiment cases in order to show the differences in configuration module for each subsection.

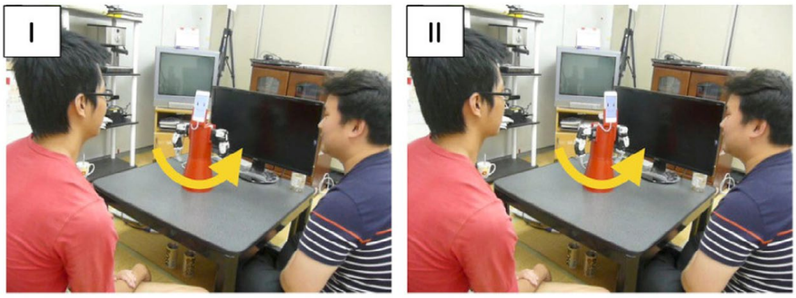


Figure 19. Experiment of Providing a Service.

7.1. Case 1: Information Service at Home

This case is on human robot interaction for elderly people who live alone. Table 8 shows what module is needed in this case 1. There is a human and a robot partner in the room (Figure 14). The robot's feelings change based on the emotional model's calculation (Figure 15). Each emotional state of the robot can be compared to the Human State (HS) number from 1 to 7 steps (Table 8). Each interaction is used as human input to calculate robot emotional state.

Experiments are performed as described in Figure 14 from I to VI. The information flow is illustrated in Figure 13. The input information consists of human detection, gesture recognition and voice recognition. The module is ready to get the utterance and gesture information from a human. When the robot receives input information, the emotion module is activated to define the robots emotion. Therefore, three modules are activated based on the emotional parameter to interact with a human: Utterance system to speech, gesture expression, and facial expression.

The detailed interaction information is shown in Table 8. In the conversation, the robot partner uses camera images and human voice information to interact with the person. In Figure 14(a)-II, the robot shows a greeting utterance to the human by using human skin-color detection information. At this stage, the robot does not track the human face. In Figure 14(a)-III, when the human expresses a hand going up and down gesture, the robot's emotion is happy based on the emotional calculation (Figure 15), because the robot does not like loneliness. In this experiment, feeling state is changed three times and mood state is changed twice (Figure 16). When positive mood is higher than negative mood, the graph shows +1 value. The opposite situation is shown as −1 value.

The robot speaks to the person by using the utterance mode based on its emotional state and the robot also could behave by using the movement parameters presented in Table 6. In Figure 14(a)-IV, the robot has sad feeling when the person says good-bye to the robot partner because then the robot will be alone. Normally, the robot selects the sentence with the higher relationship strength in the CFUS mode. However, the robot selected the "bye-bye" sentence using negative mood information since the robot has sad feeling in HS7 (Figure 17). In this way, we can realize the integrated robot partner system.

7.2. Case 2: Concierge Service at the Event Venue

This case is on information support to humans. The robot partner requires the human detection module when the robot needs to know whether human is there or not as shown in Table 9. Table 7 shows what module is needed in case 2. Input information consists of human detection and voice recognition. Figure 18 shows a human face detection situation. The robot tries to interact with the human when the robot is aware of the humans face. Therefore, robot partner can perform information support via a time dependent utterance system. In addition, according to the reaction of the human, the robot continually interacts with human. This system can also be used as information guide such as in the event venue and ticket box.

7.3. Case 3: Staff Supporting Service in Restaurants

This case is on providing a service to a human. Generally, service robots need to interact with people individually. Therefore, this robot uses its waist movement structure in order to perform a service to a person in places such as a coffee house or restaurant.

This part shows an example of a service to check whether the person needs a soft drink or not when there are two or more people as shown in Table 10. The robot tries to detect human faces with waist movement to find the human (Figure 19). Then, the robot tries to interact with the human when the robot is aware of the humans face. Table 7 shows what module is needed in this case. People who are waiting for the waiter to come can utilize these modules. Therefore, the robot partner can perform a service provision through its utterance system.

8. Conclusion

This paper introduced the integrated cognitive model of a smart device based robot partner. Modular architecture of cognitive model concepts can guide what components are needed to develop a robot partner based on the environment. This is realized by using supplementary modules that are necessary to ease the connection of the module number of features. The robot is able to understand human gestures and sentences and changes its emotional state based on the modules of system. It can produce gestural and facial expressions and verbal communication using its emotional module. The cognitive model is validated by real-world experiments. We also discuss the configuration of a minimum hardware specification and the module selection to reduce the computational cost.

As a future work, our aim is to expand the Open Source Software (OSS) of the robot partner system as a robot development platform with modular structure. We also intend to develop a robot partner system based on a smart device with the maximum capabilities of the software architecture and minimum functionality of hardware in order to reduce the costs of the robot partner. Additionally, we also consider the integration of the robot partner with the informationally structured space concept (Tang et al., 2015).

Acknowledgment

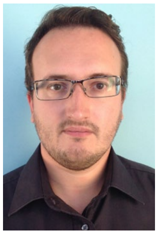
This work was partially supported by MEXT Regional Innovation Strategy Support Program: Greater Tokyo Smart QOL (Quality of Life) Technology Development Region. The authors would like to thank Peter Boyd for his proofreading.

Disclosure statement

No potential conflict of interest was reported by the authors

Notes on contributors

Jinseok Woo graduated from the Kumoh National Insitute of Techonology, Republic of Korea in 2009, and received a Master of Engineering degree from Tokyo Metropolitan University, Japan in 2011. He received his Ph.D. degree from Tokyo Metropolitan University, Japan in 2017. He is currently a research assistant professor in the Faculty of System Design, at Tokyo Metropolitan University since 2017. His current interests are in the fields of mechatronics, computational intelligence, and cognitive robotics for human-friendly robot partner.



János Botzheim earned M.Sc. and Ph.D. degrees in Computer Science at the Budapest University of Technology and Economics in 2001 and 2008, respectively. He is an associate professor in the Department of Automation at Széchenyi István University, Hungary. He is a member of several scientific societies such as John von Neumann Computer Science Society, Hungarian Fuzzy Association, and IEEE. His research interests are computational intelligence, and cognitive robotics.

Naoyuki Kubota received an M.Eng. degree from Hokkaido University, Hokkaido, Japan, in 1994, and a D.E. degree from Nagoya University, Nagoya, Japan, in 1997. He joined the Osaka Institute of Technology, Osaka, Japan, in 1997. In 2000, he joined the Department of Human and Artificial Intelligence Systems, Fukui University, as an Associate Professor. He joined the Department of Mechanical Engineering, Tokyo Metropolitan University, in 2004. He is a Professor with the Department of System Design, Tokyo Metropolitan University, Tokyo, Japan.

References

Alami, R., Chatila, R., Clodic, A., Fleury, S., Herrb, M., Montreuil, V., & Sisbot, E.A. (2006). Towards human-aware cognitive robots. In *The Fifth International Cognitive Robotics Workshop (AAAI'06)*, Boston, MA.

ARDUINO (2017). Arduino Uno. Retrieved from https://arduino.cc/en/Main/arduinoBoardUno

Apple Inc. (2011). NSLinguisticTagger. Retrieved from https://developer.apple.com/library/ios/documentation/Cocoa/Reference/NSLinguisticTagger_Class/

Apple Inc. (2016). Device Compatibility. Retrieved from https://developer.apple.com/library/content/documentation/DeviceInformation/Reference/iOSDeviceCompatibility/DeviceCompatibilityMatrix/DeviceCompatibilityMatrix.html

Bartneck, C., & Reichenbach, J. (2005). Subtle emotional expressions of synthetic characters. *The International Journal of Human-Computer Studies, 62*, 179–192.

Beck, A.T., & Haigh, E.A.P. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual review of clinical psychology, 10*, 1–24.

Botzheim, J., Woo, J., Wi, N.T.N., Kubota, N., & Yamaguchi, T. (2014). Gestural and facial communication with smart phone based robot partner using emotional model. In *Proc. of World Automation Congress (WAC2014)*, Waikoloa, HI, USA, pp. 644–649.

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., … Mulanda, D. (2004). Humanoid robots as cooperative partners for people. *International Journal of Humanoid Robotics, 1*(2), 1–34.

Burghart, C., Mikut, R., Stiefelhagen, R., Asfour, T., Holzapfel, H., Steinhaus, 70 P., Dillmann, R. (2005). A cognitive architecture for a humanoid robot: A first approach. In *Proc. of 5th IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan, pp. 357–362.

Cangelosi, A., Tikhanoff, V., Fontanari, J.F., & Hourdakis, E. (2007). Integrating language and cognition: A cognitive robotics approach. *IEEE Computational Intelligence Magazine, 2*, 65–70.

Castellano, G., Villalba, S.D., & Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In *Affective computing and intelligent interaction (ACII 2007)*, Lisbon, Portugal, pp. 71–82.

Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, A., Siméon, T. (2006). How may I serve you?: A robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, Salt Lake City, Utah, USA, pp. 172–179.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical transactions of the royal society of London B: Biological sciences, 364*, 3453–3458.

Gerstner, W., & Kistler, W.M. (2002). *Spiking neuron models*. USA: Cambridge University Press.

Google and conducted by Ipsos MediaCT (2013). Our mobile planet: Global 95 smartphone users. Retrieved from https://think.withgoogle.com/mobileplanet/en-gb/

Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.

Kurup, U., & Lebiere, C. (2012). What can cognitive architectures do for robotics? *Biologically Inspired Cognitive Architectures, 2*, 88–99.

Laban, R. (1980). *The mastery of movement*. Hampshire, UK: Dance Books.

Laban, R., & Lawrence, F.C. (1947). *Effort*. London: MacDonald and Evans.

Lourens, T., Van Berkel, R., & Barakova, E. (2010). Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Robotics and Autonomous Systems, 58*, 1256–1265.

Maass, W., & Bishop, C.M. (1999). *Pulsed neural networks*. Cambridge, Massachusetts, USA: MIT Press.

Morita, J., Nagai, Y., Moritsu, T. (2013). Relations between body motion and emotion: Analysis based on laban movement analysis. In *Proc. of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1026–1031). Berlin, Germany.

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of industrial ergonomics, 15*, 3–11.

Nishimura, K., Kubota, N., Woo, J. (2012). Design support system for emotional expression of robot partners using interactive evolutionary computation. In *Proc. of the International Conference on Fuzzy Systems*. Brisbane, Australia.

Plutchik, R. (2001). The nature of emotions. *American Scientist, 89*, 344–350.

Rane, P., Mhatre, V., & Kurup, L. (2014). Study of a home Robot: JIBO. *International Journal of Engineering Research and Technology (IJERT), 3*, pp. 490–493.

ROBOTIS (2017). DYNAMIXEL. Retrieved from http://www.robotis.us/dynamixel/

Romo (2012). Romo - The Smartphone Robot for Everyone. Retrieved from https://www.kickstarter.com/projects/peterseid/romo-the-smart-phone-robot-for-everyone/

Sakata, Y., Botzheim, J., Kubota, N. (2013). Development platform for robot partners using smart phones. In *Proc. of the 24th International Symposium on Micro-NanoMechatronics and Human Science* (pp. 233–238). Nagoya, Japan.

Schwefel, H.-P. (1981). *Numerical optimization of computer models*. New York, NY: John Wiley & Sons.

Sims, K. (1992). Interactive evolution of dynamical systems. In *Proc. of the First European Conference on Artificial Life* (pp. 171–178). Paris, France.

Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capacities of EC optimization and human evaluation. *Proceedings of the IEEE, 89*, 1275–1296.

Tang, D., Yusuf, B., Botzheim, J., Kubota, N., & Chan, C. S. (2015). A novel multimodal communication framework using robot partner for aging population. *Expert Systems with Applications, 42*, 4540–4555.

u-blox (2017). OLS426. Retrieved from http://support.connectblue.com/display/Dashboard/OLS426

United Nations. (2012). World population prospects: The 2012 revision. Retrieved from https://esa.un.org/wpp/unpp/panel_population.htm

Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation, 11*, 151–180.

Woo, J., & Kubota, N. (2013). Conversation system based on computational intelligence for robot partner using smart phone. In *Proc. of IEEE International Conference on Systems, Man, and Cybernetics, Manchester, United Kingdom* (pp. 2927–2932). Retrieved from https://doi.org/10.1109/SMC.2013.499

Woo, J., Wada, K., & Kubota, N. (2012). Robot partner system for elderly people care by using sensor network. *4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob2012)*, Rome, Italy, pp. 1329–1334.

Woo, J., Botzheim, J., & Kubota, N. (2014a). Facial and gestural expression generation for robot partners, *25th 2014 International Symposium on Micro-NanoMechatronics and Human Science (MHS 2014)*, Nagoya, Japan, pp. 218–223.

Woo, J., Botzheim, J., & Kubota, N. (2014b). Conversation system for natural communication with robot partner, *10th France - Japan Congress, 8th Europe - Asia Congress on Mecatronics (Mecatronics 2014)*, Tokyo, Japan, pp. 349–354.

Woo, J., Botzheim, J., & Kubota, N. (2015). Verbal conversation system for a socially embedded robot partner using emotional model, *Proceedings of the 24th IEEE International, Symposium on Robot and Human Interactive Communication (RO-MAN 2015)*, Kobe, Japan, pp. 37–42.

Yorita, A., Botzheim, J., & Kubota, N. (2013). Emotional models for multi-modal communication of robot partners. *2013 IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 1–6.