Virtual Emotion for Robots
– What, Why, and How –

Yutaka MIYAJI* and Ken TOMIYAMA**

* The School of Social Informatics, Aoyama Gakuin Univ., 5-10-1 Fuchinobe, Chuoku, Sagamihara, Kanagawa 252-5258 JAPAN
uta@si.aoyama.ac.jp
** Future Robotics Technology Center, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016 JAPAN
drtomiyama@gmail.com

Abstract: We propose a concept of Virtual Kansei for robots and describe our attempts to construct a series of systems for generating virtual emotion, a subset of Virtual Kansei, of a robot. Here, Kansei is a Japanese word that encompasses affection, emotion, and related mental functions. Robots need to have human-like Kansei in order for them to become true partners of humans. The proposed virtual emotion consists of three modules; emotion detector to detect emotional state of the partner, emotion generator to generate robot emotion from partner’s emotion, and expression modulator to modify robot motions to express emotion. The emotion detector utilizes facial images, voice sounds, and body motions as inputs and Bayesian networks to integrate the information from three inputs. The emotion generator is a Petri-net that has a feedback structure to represent dynamics of emotion transition. Genetic Algorithm is adopted for adjusting arc weights of the Petri-net. The expression modulator uses an emotion vector to mix basis motions corresponding to the six emotions of Ekman. Experiments and results of three modules are discussed here.

Keywords: Virtual emotion for robots, Emotion detection, Emotion generation, Emotion expression, Motion modulation

1. INTRODUCTION

As robots are coming into more personal aspects of every-day human life, they are required to interact more closely with humans. We believe robots need to have a function to have their own Kansei (a Japanese word that encompasses affection, emotion, and related mental function) in order for them to become true partners of humans. Here, we propose and discuss the concept of virtual Kansei for robots and summarize our attempts over 20 years to construct a series of systems for generating virtual emotion of a robot.

2. BACKGROUND

There are many specific fields where the introduction of robots is intensely longed for. Care and welfare field is one such field and we target it as our main application field. Japan is experiencing unprecedented rapid transformation into an aged society where the rate of aging, defined as the ratio of over 65 population versus the entire population, reached 27.3% as of October 2016 and is expected to reach 33.3% in 2036 [1]. This is the reason why the introduction of robotics into care and welfare field is an urgent matter for Japan.

We launched a project on welfare robotics in 1997 and set our objective to develop robot systems that help caretakers. Thus, we arrived at the concept of ‘caretaker support robot’ that helps caretakers by bearing those workloads which are more fit for robots and creates spare time for caretakers to interact with cared persons in such a way that only humans can do.

One of the more important tasks of the caretaker support robot is to watch the cared person and report not just physical but mental anomalies to caretakers. For this purpose, the robot needs to have a capability to find mental status of the cared person. Furthermore, the robot should be able to keep watching the cared person without irritating him/her. This requires a capability to adjust its reactive behavior to the taste of the cared person. We found out, through extensive observation of actual works of caretakers, that caretakers are matching their mental status with those of the cared person. The caretaker support robot should have this capability of caretakers, namely, the Kansei of the caretakers. Thus, a concept of Virtual Kansei for robots is proposed and studied.

3. VIRTUAL KANSEI AND VIRTUAL EMOTION

The Virtual Kansei, as its name suggests, is a function for a robot to behave like having Kansei. Our approach started from emotion, a major component of Kansei. The
emotion is a more accessible component of Kansei because it can be considered in the stimulus versus response framework, which is an input/output relationship in engineering terminology. Therefore, virtual emotion was developed first and this paper reports various attempts and outcomes regarding the virtual emotion.

### 3.1 Construction of Virtual Emotion

From our series of observations in early studies, we decided to adopt a three-module structure for our virtual emotion: emotion detector, emotion generator, and expression modulator. A conceptual robot control system that includes three modules of virtual emotion is depicted in Fig.1.

![Figure 1: The conceptual robot controller with three virtual emotion modules.](image)

It was found from observations of caretakers that they first identify emotional status of the cared person and then express the emotional state that matches the found emotion of the cared person. Following this observation, we constructed an emotion detector and an emotion generator that takes the partner’s emotion as the input.

It is often a case that we do not exhibit our true emotion but express an emotional state that matches the situation. This observation is the reason for the separate module, expression modulator, for expressing emotion. This module is responsible to express robot’s emotion by modifying ordinary motions, not by designated gestures.

### 4. EMOTION DETECTOR

The output of the emotion detector is the partner emotion. Here, the definition of six basic emotions by Ekman is adopted because of its wide applicability [2]. They are; Anger, Disgust, Fear, Happiness, Sadness, and Surprise. We added Neutral to represent non-emotional state to this list. We also decided to treat Surprise separately from other emotions due to its unique characteristics of high speed response time [3].

Emotion detector is structured as a cascade of three parts; the physical feature extraction part, individual emotion detection part, and the emotion state determination part. The input of the module can be facial images, tones of speaking, other bodily motions, or combinations of those. We adopted Bayesian network [4] for emotion detection from each physical input. The adoption of Bayesian network is motivated by the need to integrate the information from multiple sources in identifying the partner’s emotion.

#### 4.1 Three Physical Inputs

We first explain physical inputs.

- **Facial Images**
  Ekman and his group proposed Facial Action Coding System (FACS) [2] where facial expressions are clarified by Action Units (AUIDs). The facial image processing part first finds sets of various feature points using Haar patterns from consecutive still face images. The AUID detection part is responsible to identify active AUs using the sets of feature points from the facial image processing part. The developed system extracts 21 AUIDs that are used as inputs to Bayesian network.

- **Voice Sound**
  A standard feature of voice sound is adopted here. Mel-Frequency Cepstrum Coefficients (MFCC [5]) are computed from the obtained voice sound recorded at 16 bit and 16 kHz. We adopted Julius [6], an open source speech recognition engine, for this task. We decided to use 12 coefficients after a set of preliminary analysis.

- **Body Motion**
  We chose physical parameters of the head and both hands as the inputs representing body motion. Kinect for Xbox360 from Microsoft was used to acquire the positions and the accelerations in 3D, that makes a data vector 18-dimension. The speed was not adopted because our observation showed speed itself has little influence on the emotion expression.

#### 4.2 Integration of Detected Emotions

As mentioned above, Bayesian networks were used to compute the probabilities of each emotion from the above three physical inputs and the resulted probabilities are combined. This implies that the Bayesian network structure is used to integrate information at decision level rather than the feature level.
The probabilities of six emotions from three Bayesian networks are integrated using mixing ratios as in Kato et al. [4] where the integrated probabilities of emotions are given by the weighted sum of the individual probabilities. The detected emotion is the one with the highest integrated probability.

5. EMOTION GENERATOR

The target is a capability of generating human-like emotion transitions. We have tested several methods for the emotion generator, starting with neural network (NN), then recursive NN (RNN), structured NN (SNN), Petri-net with Genetic Algorithm (PN+GA), Hidden Markov Model (HMM) [7], and a hybrid NN and state space model (NN+SS).

We concentrate on the PN+GA approach here. Petri-net is a modelling tool of discrete distributed systems. It consists of Transitions, Places, Directed Arcs with weights, and Tokens [8]. The main reasons of our adoption of Petri-net are its graphical representation and ease of intentional modification of the network. The former makes it easier to understand how the internal operation pattern changes and that makes the latter possible.

It is noted that the emotion transition is controlled by a dynamics. We experience that we cannot change our emotion suddenly and that we express different emotion even if we are in the same situation. This lead us to adopt models with dynamics and introduced the concept of feedback structure.

5.1 Petri-net + Genetic Algorithm Model

When modelling a system by a Petri-net, we need to determine the network structure and to adjust the weights of the arcs. The emotion generator we developed has four layers: emotion transitions, inner places, inner transitions, and output emotion places (Fig. 2). The first layer is divided into two sets of transitions; one for the input of partner’s emotion and the other for the feedback of robot emotion at one step in the past. It is noted that we limited only one each of the input and feedback emotion transitions corresponding to the current emotion (input) and the robot emotion (feedback) is fired.

In this study, the five emotions (Anger, Disgust, Fear, Happiness, and Sadness) are divided into two levels of strong and weak emotions. By adding the Neutral emotion, this makes the total of 11 emotions and accordingly 11 input transitions and 11 output places. The number of the feedback transitions is also 11 and this makes the number of emotion transitions 22. The numbers of middle layer places and transitions are empirically determined as 22 and the corresponding places and transitions are connected by a single arc. The weights of those single arcs are set at prime numbers so that the uniqueness of the network is secured. The weights of arcs from the input transitions to the inner places and from the inner transition to output emotion places are adjusted using Genetic Algorithm so that the input and output emotion transition matches that of a human.

6. EXPRESSION MODULATOR

The naming of this module comes from the fact that it is free for a robot to express the generated emotion using many features such as light and sounds and not limited to humanly methods. However, we started our study from modulating motions of robot tasks such as handing an item to the partner. We proposed Motion Modulation System (MMS) for this purpose (Fig. 3).

The main idea of MMS is to modulate the trajectory of robot motions to achieve a given task without jeopardizing the completion of the task. We first prepared six basic trajectories that can express six basic emotions. At the same time, we devised an emotion vector generator that converts the time-series of single emotion supplied from the emotion generator into a vector of emotions. The
emotion vector makes it possible to mix six basic trajectories not only to express mixed emotions but also to give rich variety to robot motion. It is noted that the emotion vector generator can be used as a buffer between the emotion generator and the expression modulator to infuse logical restraints on expressed emotions.

6.1 Generation of Emotion Vector

The emotion vector generator actually generates a time series of five-dimensional emotion vectors excluding the neutral emotion. Each component of the vector will have an integer between 0 and 10. The neutral emotion is expressed by the all zero vector. The adopted algorithm, developed by observing natural human emotion transition, is as follows:

1. At time t, the values of all elements are reduced to half the value at time t-1. If the value is odd, say n, then (n+1)/2 is used.
2. At time t, an emotion is supplied by the emotion generator. Then the element of the vector corresponding to the input emotion will be increased by 5. If the sum exceeds 10, then the value is truncated to 10.
3. If the same emotion is not input for five times, then the corresponding element of the vector is set at 0.

6.2 Motion Parameter MP

It is assumed that the robot motion from the initial point to the final point in 3D space is represented by three cubic spline curves in x-, y- and z-planes with five notes each. The motion parameters MP adjust positions of the internal 3 nodes to modulate the motion between the initial and final points. They are:

- Duration parameter W: Changes the motion duration
- Expansion parameter Ex, Ey, and Ez: Expand trajectories along each axis
- Node position parameter R: Modify the time intervals between nodes

The motion time is prolonged (shortened) from the standard duration when W is positive (negative). The trajectory undulates when there is a mix of signs of expansion parameters. The internal nodes are evenly placed at 1/4 duration of the total time duration when the node position parameter R=0. The nodes are closer to the final (initial) node when R>0 (R<0) and, therefore, the latter (former) part of the trajectory is accelerated. It is noted that there are limits on those parameters due to the hardware limitations.

In a series of preliminary experiments, trajectories corresponding to various combinations of MPs are generated and their emotional impressions are tested for the determination of five basic trajectories for five emotions. The Neutral trajectory is the one with all 0 parameters.

7. EXPERIMENTS AND RESULTS

The experiments and results of individual modules are discussed starting from the emotion detector.

7.1 Emotion Detector

As was explained in Section 4, three physical inputs, facial images, voice sounds, and body motions are used. Separate Bayesian network is developed for each input and the results are combined at the decision level. A set of two samples of six emotions from six experiment participants is used in this experiment. The resulting Bayesian networks and the mixing of three emotion probabilities are shown in Fig. 4. The mixing ratios in Fig. 4 are computed to maximize the matching between the intended emotion and the output emotion. The result of integrating all three inputs is shown here (Table 1). Other results are referred to [9].

![Figure 4: Emotion estimation using Bayesian networks](image)

In evaluating this result, we need to consider the nature of human emotion. Through our long history of human emotion estimation studies, we found that human emotion transitions are influenced by many factors such as climate, temperature, noise, physical conditions, and so on. This fact is represented by seemingly multi-valued relationships with any set of inputs. Therefore, the estimation accuracies of over 80%, which are common in deterministic engineering relationships, are only achievable with highly controlled data. It is our
experience that those figures are out of reach with natural data and over 70% is a realistic target to achieve. From this point of view, the results shown in this table are satisfactory because the average detection rate exceeds 70%. It is noted that the ‘other subject’ raw is added to see if an emotion detector developed for a person can be used for other persons. The result shows a possibility. But we decided to develop specialized emotion detector for each partner because humans are obviously adjusting the detection method to each partner.

Table 1: Emotion estimation from all three inputs.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Joy</th>
<th>Sadness</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Neutral</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same subject</td>
<td>78.3</td>
<td>67.7</td>
<td>78.5</td>
<td>69.8</td>
<td>66.2</td>
<td>74.8</td>
<td>72.5</td>
</tr>
<tr>
<td>Other subjects</td>
<td>74.5</td>
<td>60.9</td>
<td>73.3</td>
<td>65.7</td>
<td>52.4</td>
<td>72.5</td>
<td>66.5</td>
</tr>
</tbody>
</table>

7.2 Emotion Generator

In adjusting the weights of arcs in the Petri-net model, the fitness function is given by the following formula, where the CE, P, S, Ci, and Ni are the numbers of correct answers, pairs in a set, sets, continuous correct answers from the first pair in i-th set, and correct answers, respectively.

\[
Fitness = \frac{CE}{C \times S} \quad (2)
\]

\[
CE = \sum_{i=1}^{S} \sum_{j=1}^{Ci} (N_i - j) \quad (3)
\]

Note that this fitness function can place more weight for continuous correct answers. In the evaluation experiment, 40 sets of 6 pairs each were prepared, where a set represents 30 sec time-interval and a pair of emotions is determined for each 5 sec sub-intervals.

Other parameters of GA are set as follows: the chromosome is a collection of all the adjustable weights of the arcs, both roulette and elite selection schemes, two-point crossover, and single locus mutation. The number of individuals and generations are 500 and 10,000, respectively.

The three best individuals are used for evaluation experiment. Here, a two-dimensional emotion model in Fig. 5, which is inspired by the Russel’s model [10] is introduced to evaluate differences between emotions quantitatively. There are two locations of Grief, Grief1 and Grief2. The distance from Grief is determined by the shorter of the distances computed using Grief1 and Grief2.

The error per set given below is evaluated using the same emotion data set yielded in the range of 0.8 to 2.6. Considering the fact that the maximum error of 12, this result is satisfactory.

![Two-dimensional emotion model inspired by the Russel’s model](image)

Figure 5: Two-dimensional emotion model inspired by the Russel’s model [10]. Here, s and w represent strong and weak emotions.

7.3 Individuation

We found that some inner transitions tend to fire much less than the rest of the inner transitions and that there is a correspondence between the output emotion and the less firing transitions. From this observation, we hypothesized that if we intentionally prohibit the firing of tactically chosen inner transitions then we can control the output emotion. To verify this hypothesis, we deter the firing of the inner transitions [11] and run the test with the same inputs as the evaluation experiment and we obtained the results in Table 2.

Table 2. Comparison of output emotions of Petri-nets between the original and modified ones (m) according to Table 2.

<table>
<thead>
<tr>
<th>Emotion output</th>
<th>Original (m)</th>
<th>Anger (m)</th>
<th>Disgust (m)</th>
<th>Fear (m)</th>
<th>Grief (m)</th>
<th>Joy (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>483</td>
<td>206</td>
<td>341</td>
<td>274</td>
<td>274</td>
<td>194</td>
</tr>
<tr>
<td>Anger(w)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anger(s)</td>
<td>29</td>
<td>343</td>
<td>19</td>
<td>8</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>Disgust(w)</td>
<td>6</td>
<td>7</td>
<td>160</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Disgust(s)</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Fear(w)</td>
<td>12</td>
<td>43</td>
<td>28</td>
<td>66</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Fear(s)</td>
<td>12</td>
<td>7</td>
<td>33</td>
<td>224</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Grief(w)</td>
<td>82</td>
<td>19</td>
<td>32</td>
<td>23</td>
<td>442</td>
<td>89</td>
</tr>
<tr>
<td>Grief(s)</td>
<td>10</td>
<td>8</td>
<td>28</td>
<td>36</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Joy(w)</td>
<td>107</td>
<td>116</td>
<td>84</td>
<td>109</td>
<td>19</td>
<td>222</td>
</tr>
<tr>
<td>Joy(s)</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>11</td>
<td>6</td>
<td>224</td>
</tr>
</tbody>
</table>

This clearly shows that the planned prohibition of the firing of the inner transitions gave rise to the emphasis of the corresponding emotion in the output emotion (shaded cells). This result shows the possibility of individuation of the emotion generator, namely to tailor-made the emotion generator for each partner.

7.4 Expression Modulator

The motion of handing an object to the partner with the right hand was chosen as the testing task. The right hand grabbing an object simulating a bottle of water is initially
resting at the bottom position. Then, the robot moves and extends its hand up to the shoulder height. The neutral motion, serve as the basic motion, is that of the linear motion between the initial and final point over 4 sec duration time. Strictly speaking, the motion of a hand of a robot is determined not only by its positions but also by its posture. Here, the posture is ignored for the sake of simplicity, except the initial and final postures that are fixed.

In preliminary experiments, three values for each PM (W, Ex, Ey, Ez: and R) are chosen and corresponding motions are generated. Then, 16 motions with prevailing features are selected for the main experiment. Those 16 motions were recorded, and their life-size images were shown to participants. 20 participants were asked to evaluate emotional contents of each motion by impression scores of 1 to 3. The resulting scores are analyzed using the canonical correlation analysis to find out the best combination of MPs for each emotion. As the resulting canonical correlations summarized in Table 3 [11], corresponding to the best combinations of MPs show, the motion modulation scheme MMS is promising.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>J</th>
<th>S</th>
<th>A</th>
<th>D</th>
<th>F</th>
<th>(Average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical correlation</td>
<td>0.77</td>
<td>0.85</td>
<td>0.67</td>
<td>0.82</td>
<td>0.79</td>
<td>(0.78)</td>
</tr>
</tbody>
</table>

8. CONCLUSIONS

Towards the society where robots are integrated to human society as effective supporters, we have been developing Virtual Kansei for robots. Here, we presented the summary of our efforts in developing virtual emotion, an important element of virtual Kansei, for robots. A three-module structure is proposed where each module is responsible for detection of partner’s emotional state, generating robot emotion, and expressing robot emotion. The first module, emotion detector, utilizes three physical inputs, facial images, voice sounds, and body motions to compute the probabilities of six emotional states defined by Ekman and Bayesian networks to combine those probabilities to determine the partner’s emotional state. The results of experiments show satisfactory detection rates. The second module, emotion generator, takes the partner’s emotional state as input and generates robot’s virtual emotion using a Petri-net structure with a feedback path for dynamics and GA for adapting arc weights of the Petri-net to reproduce human emotion transition. Motion modulation system (MMS) is proposed as the realization of the third module. Six motions representing six basic emotions are mixed using the emotion vector, which is constructed from the time series of robot emotion generated by the second module.

Three modules need to be integrated together and implemented on a robot to prove the validity of our approach. That is the most important task to be pursued.

ACKNOWLEDGMENTS

There are many care facilities that helped us in understanding the needs of the caretaking field and in proposing the concept of caretaker support robot. We were fortunate to have many enthusiastic students, K. Kamijo, T. Ono, T. Yanadori, J. Kogami to name a few. We acknowledge that this research was launched as a project at the Research Institute of Aoyama Gakuin University, and express our appreciation for financial as well as clerical help.

REFERENCES