

Emotional Exchange of a Socially Interactive Robot

Dong-Soo Kwon^{1,6}, Myung Jin Chung^{3,6}, Jong C. Park^{2,6}, Chang D. Yoo^{3,6}, Eun-Sook Jee^{4,6}, Kyung-Sook Park^{1,6}, Young-Min Kim^{1,6}, Hyoung-Rock Kim^{1,6}, Jong-Chan Park^{1,6}, Hye-Jin Min^{2,6}, Jeong Woo Park^{3,6}, Sungrack Yun^{3,6}, and Kyung-Won Lee^{5,6}

> ¹Dept. of Mechanical Engineering, KAIST, Daejeon, Korea ²Computer Science Division, EECS Department, KAIST, Daejeon, Korea ³Electrical Engineering Division, EECS Department, KAIST, Daejeon, Korea ⁴JS Human & Robot Music Research Institute, Daejeon, Korea ⁵Bundang Saemmul Christian School, Seongnam-si, Korea ⁶Human-Robot Interaction Research Center, KAIST, Daejeon, Korea (e-mail: kwonds@kaist.ac.kr)

Abstract: This paper presents an emotional exchange framework for a socially interactive robot. The purpose of emotional exchange in social interaction between a robot and people is to make people feel that the robot is a believable living assistant, not a mere machine for information translation. Our emotional exchange framework is composed of the emotion recognition, generation, and expression systems. A user's emotion is recognized by multi-modality such as voice, dialogue, and touch. The robot's emotion is generated according to a psychological theory about cognitive emotions caused by the social interaction within people. Furthermore, the emotion intensity is regulated by the loyalty level of a robot to various users. The generated emotion is dynamically expressed by facial expression, gesture, and the musical sound of the robot. The proposed system, which is composed of all the three components that are necessary for a full emotional interaction cycle, is implemented in the real robot system and tested. The proposed framework can be a cornerstone for the design of emotion interaction and generation systems for robots.

1. INTRODUCTION

In recent, there have been various efforts on formal approaches of how the emotions can significantly have an influence on the cognitive process of socially interactive agents (Kismet of MIT, WE-4 of Waseda Univ., Cherry of Central Florida Univ., CERO of Royal Institute of Technology, Vikia of CMU, Steward Robot Joy of KAIST, etc.) and some studies report that emotions play a role of more sophisticated communication channel in social interaction with people by externally demonstrating its internal state changes, increasing its believability by imitation of expression, and forming empathy with them(Breazeal, 2004; Hudlicka, 2003; Lisettie, 2002).

The purpose of the socially interactive robot is to coexist in a human's daily life and offer services in accordance with a human's intention. Therefore, it should be able to understand the human's emotional interaction methods and use them to communicate. In our research, human-robot emotional exchange means that a robot can transmit its own intention and internal states to people by means of emotion and also understand their affective intention by means of emotion.

First of all, the robot should be able to recognize the human's emotional states in order to fully understand their intention. Human's emotional states are expressed by facial expression, voice, language, gesture, and physiological signals; therefore, the robot can discriminate a human's emotional states with its own visual, acoustic, and tactual sensors. Secondly, the robot should be able to integrate information about humans, the environment, and itself and generate the robot's own emotional state, and reflect them by its behavior, decision making, learning, attention shift, etc. These are conducted by the robot's cognitive system for emotional appraisal and their functions are developed to be similar with the human's brain. We call it emotion generation system, which links emotion recognition with emotion expression. Thirdly, the robot should be able to express its own emotional states so that the human can also understand. This can be done by using facial expressions, gestures or sounds that are similar to the human's.

In order to integrate all the three components for emotional exchange, our emotional interaction framework was designed. In this stage, the organization of each component and each piece of information which should be exchanged between them is clarified in the home service robot domain. Through the framework design process, the necessary functionalities of each component were defined. The emotion recognition system is for understanding a human's 4 emotional states(Happiness, Anger, Sadness and Neutrality) using the human's voice, language, and touch. The emotion generation system is for generating the robot's reflective and cognitive emotional states in response to the human's emotion state and task/environmental information in the home service domain and modulating the intensity of those emotions by the different loyalty level of a robot to various users. The emotion expression system is for dynamically transmitting the robot's emotional state using an avatar's facial expression and sound effects. The proposed components for emotional exchange can be a foundation for further development and improvement of an emotion interaction system for socially interactive robots since the design was created from the topdown and holistic view-point.

In this paper, the overall emotion interaction framework is briefly explained and, the functions and technical details of each component included in the framework are described. Finally, we introduce our integrated emotional robot and discuss further works for the evaluation of our approach.

2. EMOTION INTERACTION FRAMEWORK

The emotion interaction framework is the architecture to integrate an emotion recognition, generation, and expression systems (Fig. 1). It works together with the task manager, which is responsible for the task execution of the service robot. The emotion recognition system judges the user's emotional state using the user's voice, dialogue, and touch information. Information about the user's id and emotional state, the robot's current task, and sensor information about environment are fed into emotion generation system and used for the robot's emotional state generation. Emotion generation system is composed of two layers of a reactive process and a deliberative process. The reactive process is hardwired with the raw information of the environment and generates a reflective emotional response which differs according to the robot's valence about external stimulus. The reactive emotional response takes precedence over deliberative emotional responses and is expressed outwardly in the first place. The deliberative process generates the robot's internal emotional states about user's id, user's emotional state, and task information. These emotional states are determined based on the cognitive appraisal theory that defines hierarchical emotion generation structure about the situation (Ortony et al., 1988; Sloman, 2005). The intensity of generated emotions is modulated by the loyalty level of a robot to different users. The loyalty level is determined by affective closeness between a robot and a user, which is represented by Sentiment Relation model (Kim, 2007). Emotional states are transmitted to the task manager or emotion expression system. The emotional states which are transmitted to the task manager are expressed by the robot's action coloring. Action coloring means that additional behaviors are executed between each of the robot's behaviors for a task and that those additional behaviors express the robot's emotional state efficiently while not affecting the robot's task execution. Emotional states that are unrelated with the robot's task are expressed by the robot avatar's facial expression and musical sound in an emotion expression system. This multi-modality can effectively express the robot's emotions by using dynamic model and synchronization. In the next section, all the components that construct the framework will be described.



Fig. 1. Emotion Interaction Framework

3. COMPONENTS OF EMOTION INTERACTION FRAMEWORK

3.1 Emotion Recognition using Speech Signals

Emotion recognition using speech signals is composed of two stages: feature extraction and classification. In first stage, features are extracted from speech signals. There are many features which describe the statistical characteristics of an emotion. We do not use all relevant features but select part of them to increase the recognizer performance. In second stage, we determine the emotion of the speech signal. The hidden Markov model (HMM) is used for classification. The parameter set of the HMM is estimated using a discriminative training algorithm to increase the recognizer performance.

3.1.1 Feature extraction

In emotion classification of speech signals, the popular features are pitch, fundamental frequency, energy contour, duration of silence, and mel frequency cepstral coefficients (MFCCs). We select the desirable features by Ada-Boost feature selection technique. The procedure is as follows.

Input

- 1. *N* training data, $(x_1, y_1), ..., (x_N, y_N)$ y_i is the label of feature x_i .
- 2. The emotion recognition result $h_{i,j}$ of x_i based on the feature *j*.
- 3. Weight initialization, $w_{1,i} = 1/N$.

Repeat the following procedures.

1. Compute error for all features, $\mathcal{E}_j = \sum_{i=1}^N w_{t,i} D_i$.

If $h_{i,i} = y_i$, D = 0. Otherwise D = 1.

- 2. Select the feature which has the minimum error, \mathcal{E}_t .
- 3. Update the weight, $w_{t+1,i} = w_{t,i}\beta_t^{e_i}$. If x_i is correctly classified using the selected feature, then $e_i = 1$. Otherwise $e_i = 0$.

4. Normalization,
$$W_{t+1,i} \leftarrow W_{t+1,i} / \sum_{k=1}^{N} W_{t+1,k}$$

3.1.2 Classification

In this paper, HMM is used for classification. To classify a speech signal, the parameter set of the HMM is estimated

using selected features. The most popular estimation method is the maximum likelihood (ML) estimation method. A different type of estimation, discriminative training, has better performance than ML estimation since it does not assume the statistical model is correct and the number of features is infinity. Thus, we estimate the HMM parameter set using a discriminative training. The approximated maximum mutual information (MMI) estimation (Ben-Yishai and Burstein, 2004) is used as a discriminative training. The parameter set is estimated by maximizing the following objective function,

$$F_{AMMI}(\theta) = \sum_{u=1}^{U} [\log P_{\theta}(\mathbf{O}_{u} \mid w_{u}) - \lambda \log P_{\theta}(\mathbf{O}_{u} \mid v_{u})]$$
(1)

where U is the number of training utterance and $\log P_{\theta}(\mathbf{O}_u | w_u)$ is the likelihood of u-th utterance \mathbf{O}_u given the correct transcription w_u . The 1-best recognition result of \mathbf{O}_u is denoted by v_u and λ controls the discrimination rate. By controlling λ , we obtain the desirable result.

3.2 Emotion Recognition from Natural Language Dialogues

To recognize emotions from natural language dialogues we focus on verbally expressed emotions of the users. We first analyzed drama scripts for soap opera that contain emotionally rich dialogue. We annotated each utterance based on four emotion categories such as joy, sadness, anger, and neutral, and created patterns for each category in the form of regular expressions by utilizing words and phrases from the utterances. The accuracy of the pattern-based recognition result from the unknown textual input utterances is around 56%.

We analyzed the major causes for the incorrect results. One of them is due to the difficulty in identifying the sentential type. If an utterance matched by one of the patterns was spoken in response to a request, the underlying emotion of the speaker would not be the same as the corresponding emotion with such a pattern. For example, the recognized emotion by the pattern from the first utterance of the speaker A in (Dialogue) is sadness, but the underlying emotion of the speaker is curiosity or surprise, rather than sadness.

(Dialogue)
A: What's so <u>sad</u>?
B: He was died. <u>Aren't</u> you <u>sad</u>?
A: Come on, why do you say I am sad? I am <u>not sad</u> at all.

Another cause is the presence of negatives such as 'not'. The

Another cause is the presence of negatives such as not. The patterns should not capture these sentences with negatives which modify the main verbs or adjectives in the patterns. However, if the pattern with a negative is in the question as the utterance of the speaker B in (Dialogue), we should not just regard it as improper either. Therefore, we have to consider both sentential types and negatives for more refined patterns.

For more refined patterns, we examined the types of the sentences including negatives that could be useful for recognizing certain emotions. We chose four adjectives (2 for positive sense and 2 for negative sense) and made five types of the sentential pattern. Then we collected blog contents containing those sentences using web search engines. Table 1 shows example sentences for the adjective 'pretty', and Figs. 2,3 show how many sentences are used when expressing certain emotions (anger, joy, and neutral) by types.

Table 1. Types of sentences 'pretty'.



(Words with Positive sense) (Words with negative sense) Fig. 2~3. Number of sentences expressing emotions

We found out that both declarative sentences with exclamations and interrogative sentences with negatives (1 & 2 in Fig. 2) are used frequently when expressing joy if the adjective has an originally positive sense. In addition, interrogative sentences with wh-words (4 & 5 in Fig. 2) are used more frequently than declarative sentences with negatives (3D in Fig. 2) when expressing anger-related emotions. However, such tendency with positive sense adjectives did not appear in the case of negative sense adjectives. We believe that the original negative sense of the word influences on expressing negative emotions such as anger and that the impact is especially high in the interrogative sentences with wh-words (4 in Fig. 3). We also believe that the performance could be further improved with refined patterns by incorporating this consideration.

3.3 Emotion Recognition from Touch Information

There have been several proposed methods to discriminate a human's touch with FSR (Force Sensing Resistor) sensors (Miwa, 2004; Iwata and Sugano, 2005). For a susceptive sense of the force and a single-point contact concentrating the force to a sensor, we selected the arrangement of sensors and covered the array with a bumpy pad to transmit the force induced by humans to sensors (Fig. 4). Touch patterns in about a 25cm*20cm area with just 9 FSR sensors could be measured with the proposed design. Touch patterns can be discriminated with features like sensing time and sensing force acquired from FSR sensors. "HIT" is done quickly, i.e. it means the total sensing time is short, and the maximum force of the "HIT" gives the criterion for classifying between "HIT" and "PAT". In the "PUSH" touch pattern, a contact position scarcely changes while a contact position of the "stroke" touch pattern is continuously changing. From this general character and signal datum shown in the Fig. 5, we were able to decide the criterion for determining "PUSH" and "STROKE". The criterion is whether the total sensing time of each sensor comprises the middle of the sensing time of the sensor that has sensed the minimum force among the FSR sensors. Using these features, the robot can recognize "PAT", "STROKE", "HIT", and "PUSH".



Fig. 3. (a) A construction for concentrating the touching force. (b) FSR sensor array.



Fig. 4. Signal data of each touch pattern.

3.4 Emotion Generation System

3.4.1 Reactive Layer

The architecture to generate an emotional response to external stimuli is composed of a reactive and deliberative layer. The reactive layer generates an immediate and automatic emotional response to a certain stimulus pattern. The emotional response caused by the reactive layer is tightly coupled to the stimulus itself. Basically, the role of reactive emotion is to deal with a situation which may cause a potential risk or requires user-dependent responses. In reactive layer, happiness, fear, anger, or sadness is generated by the stimulations' valence and the robot's current attention when those stimulations are different from the robot's expectation. The proposed rules for reactive emotion generation are listed in Table 2.

Table 2. Rules for reactive emotion generation.

Focus of Attention	Unexpectedness	Valence	Emotion Response
Attended	Low	Positive	Happiness
Attended	Low	Negative	Anger
Attended	High	Positive	Happiness
Attended	High	Negative	Anger
Not Attended	High	Positive	Happiness
Not Attended	High	Negative	Fear

3.4.2 Deliberative Layer

In the deliberative layer, the emotional expression and modulation of the robot's behavior is determined by the cognitive appraisal of the user's emotional state, id, previous experiences, the status of current task, and the robot's motivational status. This information is interpreted into context to be appraised in the extended OCC model, e.g. appealingness, praiseworthiness, and desirability by the context awareness module. The OCC model is the cognitive appraisal model for emotion which is proposed by Ortony et al (Ortony et al., 1988). 8 cognitive emotions are generated in the sense of reinforcement learning-based interaction (Fig. 6) (Park, Lee, Kim, and Bien, 2007).



E1 : Hope/Fear, E2 : Joy/Distress, E3 : Pride/Shame, E4 : Liking/Disliking

Fig. 5. Emotional interaction based on reinforcement learning architecture.

According to parameters which are determined in the context awareness module, the robot's deliberative emotion is determined in the extended OCC model which includes the emotional state of "embarrassment" for the situation that the above 8 emotions don't cover. The emotional state can be modulated by the loyalty module. The loyalty module (Kim et al., 2007) can change the emotional intensity according to the sentiment relation between the human and robot. In addition, the personality module can change the intensity and duration of the generated emotion by the user's preference to the robot's personality. The emotional state generated in the emotion generation system can be expressed in the novel concept of action coloring, which is proposed to overcome the monotone of robot behaviors and make HRI (Human-Robot Interaction) more natural by the abundant ability of robot expression. Action coloring is composed of pre/poststereotyped action and emotion intensity modulation. Two stereotyped actions are defined between each action step of the robot's task sequence. These do not affect the robot's actions for task execution and express emotional state effectively. The emotion with intensity can be expressed purely by facial expression, gesture, and sound. This is done in the emotion expression system.



Fig. 6. Action coloring using stereotyped action and emotion intensity modulation.

3.5 Emotion Expression using Facial Expression

Overall structure for robots' facial expression from stimuli is shown in Fig. 8 (Lee et al., 2007). It is assumed that emotion is an object moving in a D_e -dimensional linear vector space (Lee et al., 2006). The characteristics of each basis are different in this vector space. And second order differential equation is used for the emotion's dynamics like Eq. (2) (Park, Lee, and Chung, 2007).

$$M\ddot{\mathbf{e}} + C\dot{\mathbf{e}} + K\mathbf{e} = \mathbf{s} \tag{2}$$

After the emotion vector \mathbf{e} is determined by Eq. (2), it is transformed to an expression vector \mathbf{p} and then \mathbf{p} is transformed to \mathbf{p} by Eq. (3).

$$\mathbf{p}_{D_{p} \times 1} = T_{2(D_{p} \times D_{p})} T_{1(D_{p} \times D_{e})} \mathbf{e}_{(D_{e} \times 1)}, \ D_{p} \ge D_{e}$$

$$^{\gamma} \mathbf{p}_{(D_{p} \times D_{e})} = f\left(\mathbf{p}_{(D_{p} \times D_{e})}\right)$$
(3)

We developed two types of face simulator to verify our proposed method. One has 6 CPs(Control Points) and the other has 9 CPs. Fig. 9 shows the number and position of CPs for each case.Fig. 10 shows a sequence of facial expression changes for each case.



Fig. 7. Overall structure from force element (r) to facial expression (p)



Fig. 8. (a) Virtual robot model simulator with 9 CPs. (b) Actual robot face model simulator with 6 CPs.



Fig. 9. A sequence of facial expression change (Face A).

3.6. Emotion Expression using Music

In HRI, emotion expression is one of the important issues for a robot to express its own emotion for interacting with humans. Since music is an emotionally rich medium, it is one of good communication media to express emotion. To reduce ambiguity, since emotion induced from music is somewhat subjective, we composed musical sound considering effects of musical factors: Tempo, Key, Pitch, Melody, Harmony, and Rhythm. These factors are briefly summarized in to reduce ambiguity. These factors and their affections by variation are briefly summarized in (Jee et al., 2007).

There are numerous emotions and adjective words to indicate each emotion. Since we are at an early stage of research, we selected twelve emotions as a basic emotion of a robot: Happy, Sad, Joy, Anger, Like, Dislike, Hope, Fear, Pride, Shame, Distress, and Embarrassment. Figure 11 shows the composed music expressing Happy of a robot. The triplets are supported by even 4 quarter notes which demonstrate a happy feeling upon a steady atmosphere. It starts with 4-note rhythmic motive which is gradually added its member of rhythm and ends with 8 ones. It is also giving playful mood. Melodic contour shows disjunctive motion (hopping motion) and is played with a bird's singing. It is another main characteristic depicting fun and play, likewise birds love to hop and fly. In addition to that, it is played in MM. $\downarrow =160$ (Allegro means fast: 120-168), rather close to Presto in tempo. The heartbeat of children is always faster than old people. Thus, the tempo is an important factor in this music as well to give gay spirit. Our motives and explanations to composite music for Sad, Fear, and Dislike are also described in (Jee et al., 2007).



(a) Score of Happy.



(b) Upward motion of structural pitch.

Fig. 10. Composed music expressing Happy of a robot.

We performed experiment to investigate the influence of the emotion stimulated by composed music and facial expression with 20 undergraduate participants. Table 3 shows the preliminary experimental results for emotional influence. For the influence of facial expression, 60 % of the respondents reported strong Happy. On the simultaneous presentation of both facial expression and music, 80 % reported strong Happy. For Sad, Fear, and Dislike, experimental results show that music sound can improve the emotional effect of facial expression.

Table 3. Influence of the emotion expression using facialexpression and music.

	Facial Expression	Facial Expression + Music
Нарру	60 %	80 %
Sad	70 %	85 %
Fear	40 %	90 %
Dislike	45 %	85 %

For more precise emotion expression, we are now study on the relationship of emotion with different intensity and musical factors (for example, very happy emotion has higher tempo than normal happy emotion).

4. SYSTEM INTEGRATION

Our socially interactive robot is a mobile-based robot which has been developing for natural human-robot interaction by using emotional functioning such as user emotion recognition by voice and touch, and emotion generation/expression by facial expression, sound effects, and LED effects. This robot can track a person's face in around 2m distances and interact with him using various emotional modality. When the user is so far from the robot, it is able to navigate to the user for satisfying its own interaction drive. It shows an emotional reaction according to the user's emotional state and the robot's current drive to interact with the human. Currently, some pilot tests to verify effectiveness of emotion interaction are being conducted. All the three components for emotion interaction were integrated into a robot system shown in Fig. 12.



Fig. 11. Emotional interaction robot system.

5. CONCLUSIONS

The proposed emotion interaction system is designed for service robot's emotional interaction capability when it interacts with human during its task execution. Even though designed system works in limited domain, the framework can be a cornerstone for developing further emotional robot development, and the components for emotion recognition and expression can be utilized in other robot's emotion applications. The proposed emotion recognition and expression system uses multi-modality to recognize or express emotional state. This multi-modality enhances recognition and expression performance.

In the service environment, a social service robot should be able to not only express the reactive emotional responses but also proactively reflect the consequences of the previous event through the learning system. Our emotion generation framework might be utilized as a useful emotional memory for the previous event. For example, the positive memorization of a robot for the previous service consequence positively updates the affective states between a robot and a user and therefore enables the robot to activate 'pride' emotion generated in strongly positive relationship. In the other hand, the negative memorization of a robot for the previous service consequence negatively updates the sentiment relation between a robot and a user, and therefore enables the robot to activate 'shame' or 'fear' emotion to recover the relationship negatively formed.

Further important issue is the development of user evaluation method. We hope that the proposed framework might be a good performance index to measure the relative difference of the satisfaction degree the users feel for the emotional interaction with a service robot.

ACKNOWLEDGEMENT

This research was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea.

REFERENCES

Ben-Yishai, A. and D. Burshtein (2004). A discriminative training algorithm for hidden Markov models. *IEEE Trans. On Speech and Audio Processing*, vol. 12, pp. 204-217.

Breazeal, C. (2004). Function meets style: Insights from emotion theory applied to HRI, *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. **34**, No. **2**, pp. 187-194.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human-computer interaction, *International Journal of Human-Computer Studies*, Vol. **59**, pp. 1-32.

Iwata, H. and S. Sugano (2005). Human-robot-contact-state identification based on tactile recognition, *IEEE Transactions on Industrial Electronics*, Vol. **52**, No. 6, pp. 1468-1477.

Jee, E.-S., C.H. Kim, S.-Y. Park, and K.-W. Lee (2007), Composition of musical sound expressing an emotion of robot based on musical factors, *IEEE International Conference on Robot and Human Interactive Communication*, pp. 637-641.

Kim, Y.M., J.C. Park, and D.S. Kwon (2007). Behavior coordination of socially interactive robot using sentiment relation model, *IEEE International Conference on Robot and Human Interactive Communication*, pp. 1034-1039.

Lee, H.S., J.W. Park, and M.J. Chung (2006). An affect-expression space model of the face in a mascot-type robot, *IEEE-RAS Int. Conf.* on *Humanoid Robots*, pp. 412-417.

Lee, H.S., J.W. Park, and M.J. Chung (2007). A linear affectexpression space model and control points for mascot-type facial robot, *IEEE Transactions on Robotics*, to be published.

Lisetti, C.L. and P. Gmytrasiewicz (2002). Can a rational agent afford to be affectless? A formal approach, *Applied Artificial Intelligence*, Vol. **16**, No. 7-8, pp. 577-609.

Miwa, H. (2004). *Study on the mental model for humanoid robots*, Ph.D. dissertation, Dept. of Mechanical Eng., University of Waseda, Tokyo, Japan.

Ortony, A., G. L. Clore, and A. Collins (1988). *The Cognitive Structure of Emotions*, New York: Cambridge University Press.

Park, J.W., H.S. Lee, and M.J. Chung (2007). Dynamic emotion model in 3D affect space for a mascot-type facial robot, *Proceedings of Conference on Korea Intelligent Robot*, pp. 173-180.

Park, K.-H., H.-E. Lee, Y.M. Kim, and Z.Zenn Bien (2007). A steward robot for human-friendly human-machine interaction in a smart house environment, *IEEE Transactions on Robotics*, to be published.

Sloman, A., R. Chrisley, and M. Scheutz (2005). The architectural basis of affective states and processes, In J.M. Fellous and M.A. Arbib, *Who needs emotions: The brain meets the machine*, New York: Oxford University Press, pp. 203-244.