Speaker-Independent Emotion Recognition for Interstate Measuring of User Based on Separation and Rejection

Bo Seong Kim and Eun Ho Kim

***Abstract*—The purpose of this paper is to develop a speaker-independent emotion recognition system for emotional interaction between humans and robots. Recognizing human emotion from speech is one of the challenges in the field of human-robot interaction. The ability to recognize emotions from an unspecified human, called speaker-independent emotion recognition, is important for commercial use in speech emotion recognition systems. However, generally, speaker-independent systems show lower performance compared with speaker-dependent systems, as emotional feature values depend on the speaker and his/her gender. Hence, this paper describes the realization of speaker-independent emotion recognition based on separation and rejection to make the emotion recognition system accurate and stable. Through comparison of the proposed methods with conventional method, the improvement and effectiveness of proposed methods were clearly confirmed.**

***Index Terms*—Speech emotion recognition, confidence measure, SID system.**

I. Introduction

Nowadays, Human-Robot Interaction (HRI) as well as Human-Computer Interaction (HCI) are promising areas in intelligent robot development and artificial intelligent [1], [2]. Unlike from industrial robots, intelligent robots will generally be used by people who are not familiar with robots. Therefore, it is necessary for a personal robot to communicate naturally with humans and reprogram itself based on that communication, and HRI should consisted of conventional and instinctual ways of accomplishing these goals. There are several ways in which humans interact with each other such as speech, eye contact, and gestures. Among them, speech communications is the most common in human-to-human interaction and also the most effective way of communication through which people can readily exchange information without the need for any other tool [3]. Therefore, many researchers have studied technologies related to speech communication. In particular, emotional interaction has become one of the most important research areas [4], as emotional interaction can help a robot to be more flexible in complex and uncertain environments [5]. For example, when a robot interacts with a human, the robot can generate suitable behaviors or express suitable emotions according to the emotional state of the human. Hence, this paper describes speech emotion recognition for emotional interaction between humans and robots.

The major challenge of speech emotion recognition for commercial use is emotion recognition from unspecified humans, emotion recognition from spontaneous speech, and emotion recognition in a real-world environment such as noisy environment [6]. In particular, speaker-independent (SID) emotion recognition has to be accomplished in pursuance of interaction with various users. However, speaker-independent emotion recognition systems show a lower accuracy rate compared with speaker-dependent systems as spoken sentences or continuous dialogue generally change based on the emotional state of the speaker and are also affected by the individual characteristics of the speaker. Hence, this accomplishment of a speaker-independent system would be a major breakthrough.

To overcoming the variability of the speaker, a novel speaker-independent emotional feature, a ratio of a spectral flatness measure to a spectral center, was suggested by Kim *et al.* [7]. In 2015, Maxim Sidorov [8], [9] proposed a novel method on speech-based adaptive emotion recognition through addition of speaker specific information and achieved 10% accuracy improvement. And Iliou and Anagnostopoulos reported around a 51% recognition rate for seven emotions using neural networks [10]. In 2011, C. N. Anagnostopoulos *et al.* [11] addressed three important issues in speech emotion recognition: emotional feature of speech, classification schemes, and an emotional speech database. From the survey of speech emotion classification, they conclude that the average recognition accuracy of a speaker-independent speech emotion recognition system is around 60%, less than 80% in most of the proposed techniques, and in some case it is as low as 50% [12].

The purpose of this study is to design a speaker-independent emotion recognition system for emotional interaction robot. To develop emotional interaction in which robots show stable and homogeneity reactions, this study proposes an emotion recognition system separately from consonant and obstruents, as well as rejection algorithms based on a confidence measure. Described within the structure of this paper is an overview of the target system, and two strategies are described for developing an emotion recognition system that is robust in speaker variation. Finally, the database used in this study as well as the method of developing the database, the experimental condition, and experimental results in the speaker-independent system are given, as is the conclusion at the end of the paper.

II. Target System

In order to promote emotional interactions between humans and robots, an emotional interaction robot, Mung, was developed with the delicate interaction response of "bruises and complexion color due to emotional stimuli"

Manuscript received January 17, 2018; revised March 12, 2018. This work was supported in part by the Korea Institute of Industrial Technology.

The authors are with the Korea Institute of Industrial Technology, Ansan-si, Gyeongi-do, South Korea (e-mail: kimeunho@kitech.re.kr).

doi: 10.18178/ijmlc.2018.8.2.679

[13]. As humans become black and blue when being physically hurt, the robot also becomes bruised when its feelings are emotionally hurt. The robot becomes neutral and blushes in response to positive stimuli (see Fig. 1). The developed robot can help humans to see their and/or others' emotional states; therefore, humans can try to change their attitudes and promote human-human relationships.

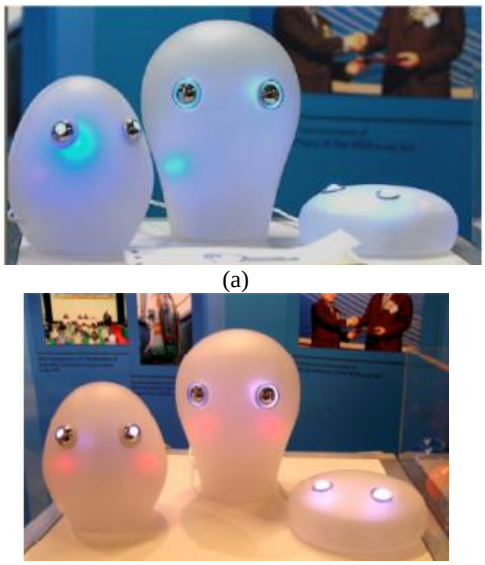

Fig. 1. An emotional interactive robot, Mung, exhibited in the 51st Annual IAEA General Conference: a) bruised robots and b) blushing robots.

The robot consists of three modules: a perception module, a robotic emotional state module, and an expression module. To represent a dynamic emotional state (short-term phenomenon) and mood (long-term phenomenon), the robotic emotional state is modeled using a mass-spring-damper system with an elastic-hysteresis spring. The personality of the robot was easily modeled by changing the coefficient of the system. In the expression module, bruises and complexion colors are expressed through full color light-emitting diodes (LEDs). The full color LEDs are controlled using pulse width modulation (PWM) signals, and the duty ratio of the PWM signal is controlled according to the emotional state of the robot to control the colors and brightness of full color LEDs.

III. Design of a Speaker Independent Emotion Recognition System

A. Strategies

In this study, two strategies are proposed to improve the performance of speaker-independent emotion recognition systems. One is recognizing the emotion separately from obstruent sounds and sonorant sounds (called sound type), as obstruent sounds and sonorant sounds are different in terms of spectral characteristics (see Fig. 2). As obstruent sounds and sonorant sounds have different spectral characteristics, most emotional features, such as linear prediction coefficients (LPCs), mel frequence cepstral coefficients (MFCCs), log frequency power coefficients (LFPCs), and energy also have different values according to sound type (see Table I).

Table II shows the result of a paired sample T test. The hypothesis that obstruent sounds have the same emotional feature value as sonorant sounds is rejected, as shown in Table II. Consequently, it is verified that obstruent sounds and sonorant sounds have different emotional feature values. If emotional features vary with sound type rather than emotions, emotion recognition can easily fail because emotional feature values changed by emotions are hidden by the changes according to sound type.

TABLE I: C Mean and Standard Deviation (STD.) of RSS for Six Speake4yrs: 'M': Male, 'F': Female

Mean(std.)	Neutrality	Sadness	Joy	Anger
M1	2.54(0.08)	2.80(0.08)	2.57(0.11)	2.37(0.09)
M2	2.54(0.12)	2.80(0.07)	2.65(0.07)	2.32(0.12)
M3	2.55(0.09)	2.68(0.09)	2.45(0.12)	2.39(0.08)
F1	2.48(0.07)	2.59(0.06)	2.40(0.09)	2.33(0.08)
F2	2.46(0.09)	2.64(0.07)	2.32(0.08)	2.24(0.09)
F3	2.56(0.07)	2.69(0.06)	2.50(0.08)	2.41(0.10)

TABLE II: The Result of Paired Sample Test

Feature	*Different mean*	*T-statistic*	*Significant level*
1st LPC	*0.35*	*69.3*	< .001
1st MFCC	*3.47*	*197.2*	< .001
1st LFPC	*0.77*	*39.4*	< .001
energy	*0.032*	*4.12*	< .001

Hence, it is necessary to recognize emotion separately from obstruent sounds and sonorant sounds.

The other strategy proposed in this paper is rejecting uncertain recognition results based on confidence level. Generally, an emotion recognition system perceives the human emotion for every sentence. However, it is unnecessary to recognize emotion at every moment during the interaction or communication. As shown in Fig. 3, during human-robot interaction, the proposed recognition system filters out unreliable recognized results and only reliable results are accepted as a final decision. As the rejection method based on the confidence level is widely used in the field of pattern recognition [14]-[16], in this study, a confidence measure for speech emotion recognition is proposed and a rejection algorithm based on confidence level is also proposed for stable and reliable speaker-independent emotion recognition.

B. Separation of Obstruent/Sonornat Sounds

Since speech signals are not stationary even in a general sense, it is common in speech emotion recognition to divide a speech signal into small frames. The signal within each frame is considered to be approximately stationary.

To separate speech into obstruent and sonorant segments, each frame is separated according to whether it is obstruent or sonorant using the spectral center, as shown in (1) where f_i is the frequency and E_i is the spectral power. After classifying each frame into sound types, a succession of identical sound types is combined, as a segment as shown in Fig. 4. Finally, the emotion is decided as shown in (2). P_{seg} is the posterior probability for each segment, W_{seg} is the weight parameter proportional to the length of the segment, and α is the weight for the sonorant segment.

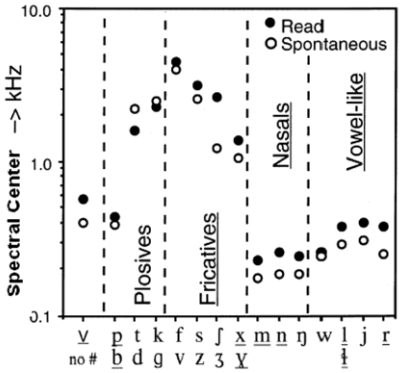

Fig. 2. Spectral center for each phoneme.

$$\text{Spectral Center} = \frac{\sum f_i \bullet E_i}{\sum E_i} \tag{1}$$

$$\text{result} = \arg\max_{e=N,A,J,S} \begin{bmatrix} \alpha \times \sum_{seg \in sonorant} w_{seg} P_{seg}^{e} \Big)_{sonorant} \\ +(1-\alpha) \sum_{seg \in obstruent} w_{seg} P_{seg}^{e} \Big)_{obstruent} \end{bmatrix} \tag{2}$$

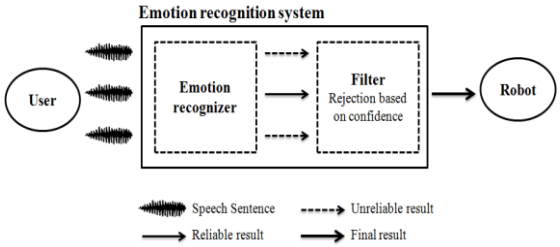

Fig. 3. Emotion recognition system with the rejection based on confidence level.

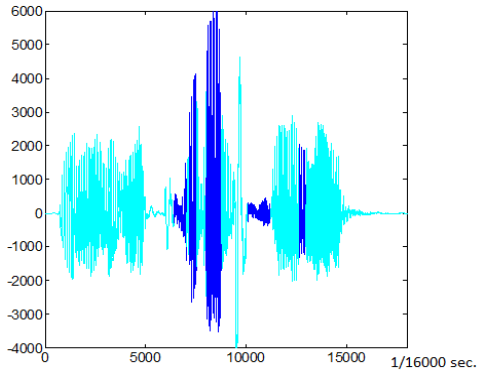

Fig. 4. An example of segmentation result: five obstruent (blue) and six sonorant segments (sky-blue).

C. Confidence Measure

To measure the confidence level of the recognition result, a confidence measure based on conditional probability is proposed in this study, as shown in (3). The confidence measure, $P((c|\vec{s}, e_i)$ is a conditional probability in which a recognition result is correct given a recognition result ($E = e_i$) and sequence of segment features (S). Based on the Baye rule and an independent assumption of the emotional state and sequence of segment features, conditional probability is reformulated, as shown in (3). R represents the recognition result, where c is the correct result and w is the wrong result. E represents emotion where e_i is neutrality, joy, sadness, and anger in this study.

To represent reliability of the recognition result of perceived speech, the normalized likelihood for each emotion and recognition consistency was used for segment features ($\vec{s} \in R^5$), as shown in (4). LH_i is the normalized likelihood of each emotion, and C is the recognition consistency. C is one when the recognition result of the segment and recognition result of the whole speech show identical results.

$$\begin{aligned} P\left(R=c|S=[s_1,\cdots s_n], E=e_i\right) &= P\left(c|\vec{s},e_i\right) \\ &= \frac{P\left(\vec{s}|c,e_i\right)\bullet P\left(c|e_i\right)}{P\left(\vec{s}|c,e_i\right)\bullet P\left(c|e_i\right)+P\left(\vec{s}|w,e_i\right)\bullet P\left(w|e_i\right)} \\ &= \frac{P\left(\vec{s}|c\right)\bullet P\left(c|e_i\right)}{P\left(\vec{s}|c\right)\bullet P\left(c|e_i\right)+P\left(\vec{s}|w\right)\bullet P\left(w|e_i\right)} \end{aligned} \tag{3}$$

$$\vec{s} = [LH_N, LH_J, LH_S, LH_A, C] \tag{4}$$

For example, a sentence consists of 11 segments (6 sonorant segments and 5 obstruent segments), and the sequence of the segment feature is as shown in (5). As the recognition result of the sentence is anger and the recognition result of first segment is sadness, the result consistency of the first segment is 0.

$$\begin{aligned} S &= \left[\vec{s}_1, \vec{s}_2, \cdots, \vec{s}_{11}\right], \quad where \\ \vec{s}_1 &= [0.20, 0.23, 0.38, 0.19, 0], \\ \vec{s}_2 &= [0.21, 0.25, 0.23, 0.31, 1], \\ &\cdots, \\ \vec{s}_{11} &= [0.20, 0.31, 0.13, 0.36, 1] \end{aligned} \tag{5}$$

D. Algorithm for SID Emotion Recognition System

The procedure of proposed speaker-independent emotion recognition system is shown in Fig. 5. The procedure is as follows:

1) Divide a sentence into N frames using a hamming window.
2) Separate each frame by sound types (obstruent/sonorant sounds) using the spectral center.
3) Make a succession of identical sound types to segments (see Fig. 4).
4) Estimate the likelihood for each emotion using classifiers trained by each segment.
5) Make segment feature ($\vec{s} \in R^5$) using normalized likelihood and recognition consistency for each sound type, as shown in (4).

6) Decide the emotion using the maximum a posterior method, see (2).a
7) Calculate the confidence measure, $P((c|\vec{s}, e_i)$..
8) Decide the acceptance of the recognition result in step 6.

IV. Experiments

A. Emotional Speech Database

Speech emotion recognition accuracy depends on the number and kind of emotions, and the type of emotion elicitation (acted versus non-acted data). The Korean emotional speech (KES) database [17] produced by Yonsei University's Media and Communication Signal Processing Laboratory is used in this study. The KES is an acted database that was collected from amateur actors and actresses. An emotional speech was acted by amateur actors and actresses who had practiced emotional expression; they were selected according to their emotional expression ability. The KES covers the four emotions of neutrality, joy, sadness, and anger. The KES contains short, medium, and long sentences that are from 0.5 to 3 seconds in length and are context-independent. The sentences were chosen considering the following criteria:

1) Easy pronunciation in neutral, joyful, sad, and angry states,
2) Natural expression of emotion from the sentences,
3) Equal inclusion of all phonemes of the Korean language,
4) Consideration of various modes of expression.

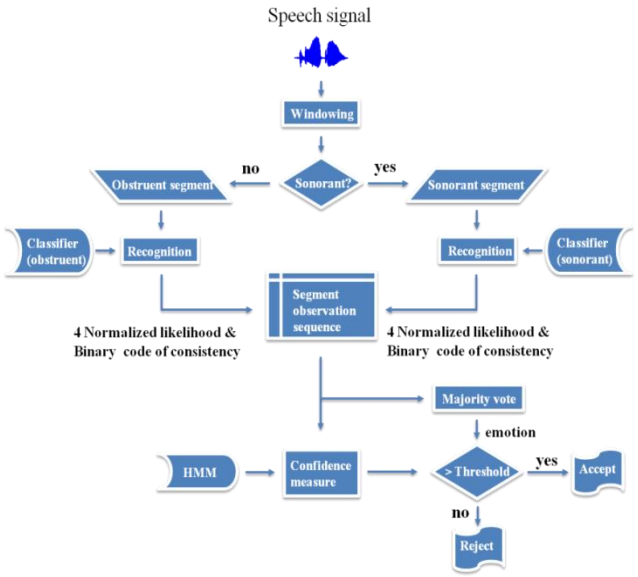

Fig. 5. Procedure of proposed SID emotion recognition system.

The KES database consists of 16,200 samples (45 dialogic sentences repeated three times in the four emotions by each of the 30 speakers who comprised of 15 males and 15 females) or 135 samples per speaker and per emotion. All sentences were recorded four times, and the recording judged the worst among the four was discarded. The aim of this step was to filter clumsy wording and maintain the consistency of the speech. Table 3 shows the human recognition accuracy of KES. On average, human subjects can determine the emotion of the KES recorded utterances with about 80% accuracy. Identical sets of sentences were used for all emotions, and the recordings were made in a silent experimental environment. The original data was stored in the form of 32-bit, 16 kHz sound with over 30dB S/N. It was margined with silence for approximately 50 ms at the beginning and end of each utterance.

B. Experimental Condition

The simulation of speaker-independent emotion recognition system was evaluated using MATLAB run on a PC. Every experiment was performed in off-line mode using the database. Thus, noise reduction from the microphone is outside of the focus of this paper.

TABLE III: Confusion Matrix of Uman Recognition Accuracy for KES Database

% Recog.	Human performance (%)			
	Neutrality	Joy	Sadness	Anger
Neutrality	83.9	3.1	8.9	4.1
Joy	26.6	57.8	3.5	12.0
Sadness	6.4	0.6	92.2	0.8
Anger	15.1	5.	1.0	78.5
Average	78.1			

As a small feature vector provides a better generalization performance [18] and it is computationally cheaper to compute lower dimensional features, in this experiment, an orthogonal-linear discriminant analysis (OLDA) was employed [19]. The OLDA computes projections that maximize the Fisher criterion and, at the same time, are pair-wise orthogonally. The method used in OLDA combines the eigenvalue solution of and the Gram-Schmidt orthonormalization procedure. S_b is the between scattering matrix, and S_w is within the scattering matrix, as defined in (6);

$$\begin{aligned} &\text{set of features} \quad y_1, y_2, \ldots, y_n \\ &c : \#\text{ of class} \\ &n_i : \#\text{ of features} \in \text{class}_i \\ &\tilde{m}_i = \frac{1}{n_i}\sum_{y\in y_i} y \quad \text{and} \quad \tilde{m} = \frac{1}{n}\sum_{i=1}^{c} n_i \tilde{m}_i \\ &S_w = \sum_{i=1}^{c}\sum_{y\in y_i}(y-\tilde{m}_i)(y-\tilde{m}_i)^T \text{ and } \quad S_b = \sum(\tilde{m}_i-\tilde{m})(\tilde{m}_i-\tilde{m})^T \end{aligned} \tag{6}$$

For the classification, the Gaussian mixture model (GMM) was employed. To estimate the likelihood, $P((\vec{s}|c)$, for confidence measure in (3), hidden Markov models with 5 states and 20 mixtures was used. To evaluate the proposed system, the leave-one-speaker-out cross-validation method was used.

C. Experimental Results

To verify the effectiveness of obstruent/sonorant sound separation, a controlled experiment was conducted using various emotional features, LPCs, MFCCs, Perceptual Linear Predictive (PLP) analysis, LFPCs, and energy. Table 4 shows a comparison between the conventional approach and obstruent/sonorant sound separation approach for each emotional feature, which is the average recognition rate, number of improvement speakers among 30, and the maximum/minimum improvement.

Among the five emotional features, MFCCs show the best performance and emotional feature of energy shows worst. On average, all features show improvement from 6.9% to 27.6% after obstruent/sonorant sounds separation.

TABLE IV: COMPARISON OF EMOTION RECOGNITION PERFORMANCE BETWEEN THE CONVENTIONAL METHOD AND SEPARATION METHOD

Feature	Separation	Average recognition rate (%)	number of improvement subjects	Max./Min. improvement (%)
LPCs	None	50.2	-	-
	Apply	60.9	26	31.5/-8.3
MFCCs	None	59.8	-	-
	Apply	72.7	27	28.3/-10.9
PLP	None	56.3	-	-
	Apply	67.1	28	41.1/-8.3
LFPCs	None	57.5	-	-
	Apply	64.4	24	23.1/-9.1
Energy	None	40.6	-	-
	Apply	68.2	30	48.5/2.6

The paired T-test was applied to compare recognition performances that are subjected to different methods (separation method versus non-separation method). The result shows that the separation method is better than the non-separation method at a 99% significance level.

To verify the effectiveness of the rejection algorithm, the 12 orders of MFCCs that show the best performance in the previous experiment were used among five emotional features. The features were computed over a hamming window with duration of 20 ms. MFCCs were computed as in (7), where M is the number of cepstrum coefficients and represents the log-energy output of the k filter. The mean value of each order was used to construct the static feature.

$$MFCC_i = \sum_{k=1}^{20} X_k \cos\left[i(k-\tfrac{1}{2})\tfrac{\pi}{20}\right], \quad i = 1,2,\ldots,M \tag{7}$$

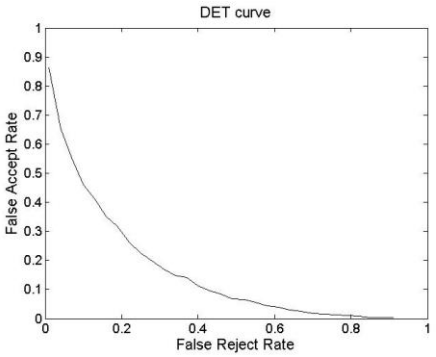

Fig. 6. Detection and trade-off curve of proposed rejection algorithms using a confidence measure; average for 30 speakers.

Fig. 6 is the detection and trade-off curve which shows the performance of the proposed confidence measure. Fig. 7 shows the average recognition rate for 30 speakers when unreliable results are rejected. As shown in Fig. 7, the recognition rate is increase as the rejection rate increases. Specifically, when a 30% recognition result was rejected, the recognition rate increased from 73% to 86% accuracy. And when half of the recognition result was rejected, the recognition rate increased to 92% accuracy.

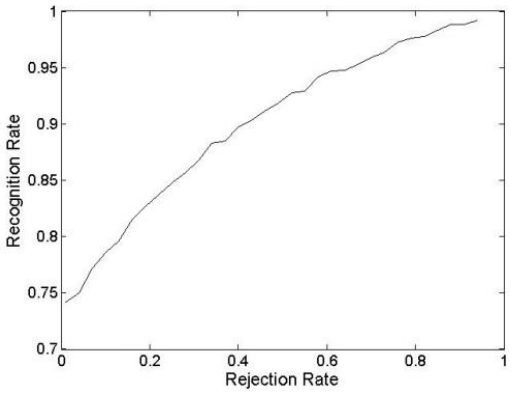

Fig. 7. Average speech emotion recognition performances for 30 speakers versus rejection rate.

As a result, the two proposed strategies which are separating obstruent/sonorant sounds and rejecting unreliable recognition results based on confidence measure are verified. This result shows the possibility of commercial use of the proposed speaker-independent emotion recognition system as the recognition rate increases from 60% to 92% accuracy.

V. CONCLUSION

In this paper, speech emotion recognition in speaker-independent systems for emotional interaction were investigated. To make a speaker-independent emotion recognition system robust and accurate, two strategies were proposed. First, based on the different vocalization of consonants and obstruents, speech emotion recognition separately from consonants and obstruents was proposed to reduce the text-dependency. Second, to make the emotion recognition system homogenous and accurate, a rejection algorithm based on a confidence measure is proposed. To decide whether the result is reliable or not, the confidence measure, which is the conditional probability that the recognition result is correct given the recognition result and sequence of segment features, is proposed.

From comparison of the proposed methods with the conventional method, the improvement and effectiveness of proposed methods was clearly confirmed. As a result, the separation algorithm is effective in improving the recognition rate from 6.9% to 27.6% for various emotional features. And when we reject half of the results based on the proposed confidence measure, the recognition rate increase from 73% to 92% in the case of MFCCs. This results show the possibility of the commercial use of the speaker-independent emotion recognition system.

REFERENCES

[1] P. A. Lasota, G. F. Rossano, and J. A. Shah, "Toward safe close-proximity human-robot interaction with standard industrial robots," in *Proc. IEEE International Conference on Automation Science and Engineering*, Taipei, Aug. 2014, pp. 273-285.

[2] T. B. Sheridan, "Human-robot interaction; status and challenges," *The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 4, pp. 525-531, 2016.

[3] A. Abelin, and J. Allwood, "Cross linguistic interpretation of emotional prosody," in *Proc. the ISCA Workshop on Speech and Emotion*, Vancouver, 2000, pp. 110-113.

[4] M. Pantic and L. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," in *Proc. the IEEE*, Nice, Sep. 2003, pp. 1370-1390.
[5] C. Breazeal, "Sociable machines: expressive social exchange between humans and robots," PhD thesis, Massachusetts Institute of Technology, Combridge, USA, 2000.
[6] B. Schuller, D. Seppi, A. Batliner, A. Maier, and S. Steidl, "Towards more reality in the recognition of emotional speech," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal*, Honolulu, HA, USA, 2007, pp. 941-944.
[7] E. H. Kim, K. H. Hyun, S. H. Kim, and Y. K. Kwak, "Improved emotion recognition with a novel speaker-independent feature," *IEEE/ASME Trans. on Mechatronics*, vol. 14, no. 3, pp. 317-325, 2009.
[8] M Sidorov, S. Ultes, and A. Schmitt, "Emotions are a personal thing: Towards speaker-adaptive emotion recognition," in *Proc. ICASSP*, Florence, May 2014, pp. 4836-4840.
[9] C. H. Wu, J. C. Lin, and W. L. Wei, "Acoustic emotion recognition two ways of features selection based on self-adaptive multi-objective genetic algorithm," in *Proc. ICINCO 2014*, Vienna, 2014, pp. 851-855.
[10] T. Iliou and C. N. Anagnostopoulos, "Statistical evaluation of speech features for emotion recognition," in *Proc. Fourth International Conference on Digital Telecommunications*, Colmar, France, July 2009, pp. 121-126.
[11] C. N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: A survey from 2000 to 2011," *Artificial Intelligence Review*, vol. 43, no. 2, pp. 155-177, 2015.
[12] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," *Neural Comput.*, vol. 9, no.4, pp. 290-296, Appl. 9, 2000.
[13] E. H. Kim, S. S. Kwak, K. H. Hyun, S. H. Kim, and Y. K. Kwak, "Design and development of an emotional interaction robot, 'Mung'," *Advanced Robotics*, vol. 23, no. 6, pp. 767-784, 2009.
[14] S. Marukatat, T. Artieres, P. Gallinari, B. Dorizzi, and P. Lip, "Batch-adaptive rejection threshold estimation with application to OCR post-processing," *Expert System with Applications*, vol. 42, no. 21, pp.8111-8122, 2015.
[15] G. Bouwman, L. Boves, and J. Koolwaaij, "Weighting phone confidence measures for automatic speech recognition," presented at Workshop on Voice Operated Telecom Services, Ghent, Belgium, 2000.
[16] H. Jiang, "Confidence measures for speech ecognition: A survey," *Speech Communication*, vol. 45, no. 4, pp. 455-470, 2005.
[17] B.-S. Kang, C.-H. Han, S.-T. Lee, D.-H. Youn, and C.-Y. Lee, "Speaker independent emotion recognition using speech signals," in *Proc. ICSLP-2000*, Beijing, October 16-20, 2000, pp. 383-386.
[18] I. Witten and E. Frank, *Data Mining*, Morgan Kaufflan, Los Altos, CA, 2000.
[19] T. Okada, and S. Tomita, "Optimal orthogonal system for discriminant analysis," *Pattern Recognition*, vol. 18, no. 2, pp. 139-144, 1985.

Bo Seong Kim received his BS degree from Mechatronics Engineering in the Korea Polytechnic University, Korea, in 2013 and received his MS degree from Dept. of Computer and Software in the Graduate School of Hanyang University, Korea, in 2017. His research interest include speech emotion recognition, emotional interactive robot, pattern recognition for Human-Robot interaction and linear/nonlinear control system, electronics circuit, and robust sensing system for robotics.

Eun Ho Kim received his B.S. and Ph.D. degree from the Division of Mechanical Engineering, School of Mechanical, Aerospace and System Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2004 and 2009, respectively. He is a principal researcher in KITECH and adjunct professor with University of Science and Technology (UST). His research interest includes speech emotion recognition, emotional interactive robot, machine learning, and pattern recognition for human-robot interaction and linear/nonlinear control system, system dynamics, and robust sensing system for robotics.