Personalized Emotion Recognition Using PLP Analysis of Speech Signals

Soojin Lee, Seong-Pyo Cheon, Gyeong-dong Baek and Sungshin Kim

School of Electrical and Computer Engineering, Pusan National University, Busan, Korea

Pusan National University 30 Jangjeon-dong, Keumjeong-ku, Busan 609-735, Korea
E-mail: pooh3887@pusan.ac.kr, buzz74@pusan.ac.kr, gdbaek@pusan.ac.kr, sskim@pusan.ac.kr

Abstract – This paper presents an emotion recognition using personalized speech signals. In the field of development of Human Interface technology, the interaction between human and machine is important. The research about emotion recognition helps this interaction. We are trying to analyze an emotion using personalized speech signals. We have to extract the feature of speech signal for emotion recognition using speech signal. This paper proposed an emotion recognition using speech signal by individual. Using Perceptual Linear Prediction (PLP), we extracted the feature of speech signal. Then, we proposed an algorithm that can evaluate the emotion from speech signal easily and real time using personalized emotion patterns that is made by PLP analysis.

Keywords: Speech signal, Emotion recognition, Perceptual Linear Prediction (PLP)

1 Introduction

Human's sound is used for communication with emotion as well as conversation. Emotion which is included at sound shows speaker's psychology and does communication with other person more naturally. In the field of emotion recognition using speech signal, we can use the method using only meaning of words or rhythmical information and the method using both meaning of words and rhythmical information. There are many studies about these methods.[1]

In the future, an intersection with human is an essential element in the development of human interface. An essential element in the intersection with human and machine is an emotion recognition method. Emotion recognition method to human is trying to propose the method to use image, speech and biologic signal[2]. J. Nicholson made feature vector using power, pitch and linear predictive coding analysis of speech and classified it into several emotions by neural network (NN). They achieved 50% accuracy for eight emotions [3]. Aishah Abdul Razak compared fuzzy model with NN for emotion recognition. They used energy, LPC coefficients, duration, pitch and jitter for feature of speech signal. They achieved about 60% accuracy for six emotions [4]. Kang, Myun-Gu used an algorithm using Vector Quantization (VQ) and Gaussian Mixture Model (GMM) for individual emotion recognition of human and sentence [5]. Cho, Yoon-Ho proposed a speech emotion recognition system that could recognize human emotional condition from real time speech signal via cellular phone. They achieved 86.5% for two emotions [6]. Vladimir Hozjan and Zdravko Kacic used individual and various languages for emotion recognition study using speech signal. In the study, it was obtained high recognition rate from dependent emotion recognition and emotion recognition of each language was obtained similar recognition rate [7].

In this paper, we proposed an algorithm to recognize the emotion using PLP analysis. PLP analysis is usually applied to decide language recognition or speaker recognition by speech signal. We proposed it for personalized emotion recognition of speech signal.

2 Speech Signal

2.1 Description of Speech Signal
Human words have much information. The purpose of the sound is usually to communicate with other people. We could keep the quantitative values of the sound using a microphone. These quantitative values are speech signal. Figure 1 shows the graph of the speech signal.

2.2 Feature of Speech Signal
There are many methods to analyze the feature of speech signal. The feature analysis of speech signal is basically classified into pitch features and energy features. The pitch feature analysis is method which makes analysis of the maximum value or minimum value of speech signal. The energy feature analysis is method which makes analysis of the time range or frequency range of speech signal. This paper proposed the emotion recognition by the PLP analysis.

PLP analysis is speech signal analysis which is concerned about an auditory sense of speech signal.

Human's auditory sense is shown in figure 2. Human's auditory sense has difference sense when they heard their ears through the frequency. PLP analysis is regulated the degree of frequency by auditory sense and classified into the spectrum analysis of speech signal by disassemble.
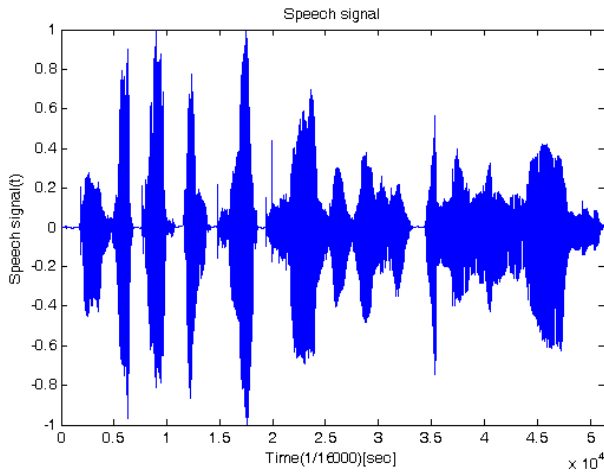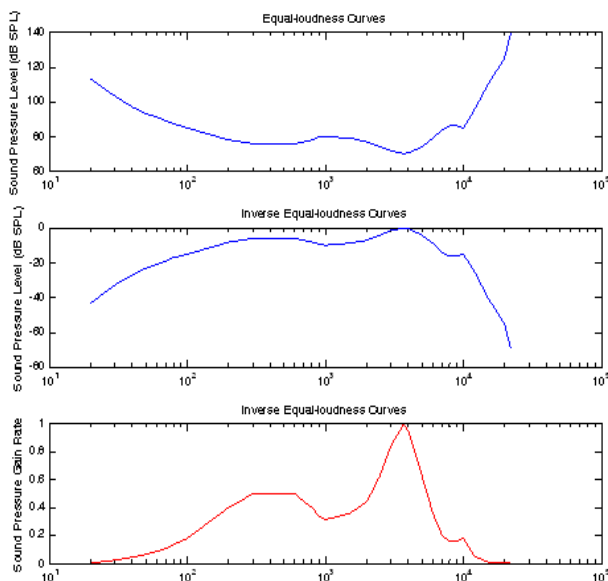

Fig. 1 Regulated speech signal.

Fig. 2 Human's auditory sense.

PLP analysis algorithm is shown in figure 3. Its Critical Band Analysis is input signal power which is changed frequency range as figure 3. Equal Loudness is expressed the strength of tone by changing frequency of emotion. PLP analysis has the feature of accent, timbre and language of speech signal.

We could get the result as a spectrum form after analyze the speech signal using PLP analysis that is shown in figure 4. A vertical axis means a frame of spectrum at the third of the figure 4. It is consist of 21 frames. In this paper, we extracted the feature of speech signal using the average energy and the Min-Max values each frame.
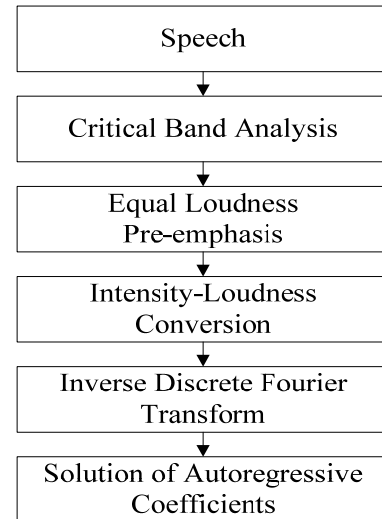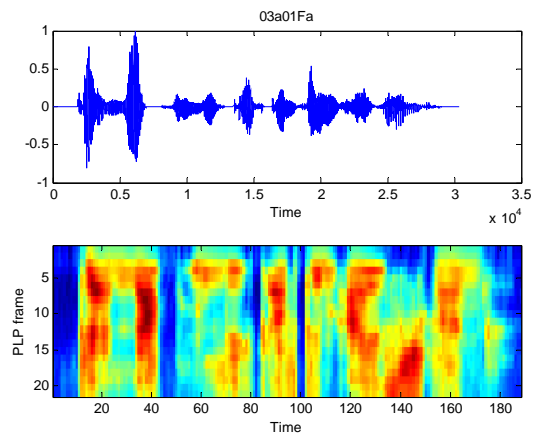

Fig. 3 The PLP analysis algorithm.

Fig. 4 PLP analysis graph of speech signal.

3 Emotion Recognition

3.1 Description of Emotion

Human expressed their emotion using behavior, expression or speech. Human emotion is essential factor for communication. That is, it can be easy transmitted what someone want when he expresses his emotion. This paper proposed an algorithm to classify human emotion into seven kinds that seen in figure 5.
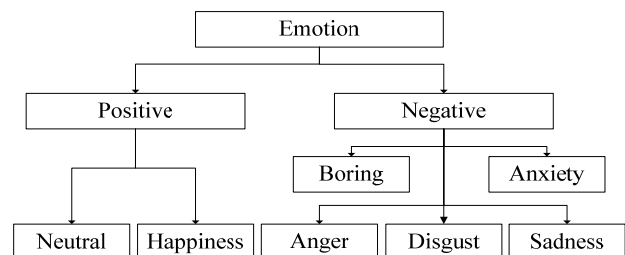

Fig. 5 The signal analysis about 7 emotions

3.2 Feature of Emotion

In order to analyze the feature of the speech signal, we regulated the speech signal through a microphone.

Basically we could analyze the speech signal in the time space. We also could analyze the speech signal in the frequency space. When we analyze the speech signal in the frequency space, we can see the change by emotion. The figure 6 shows the graph of the time range and frequency band when speaker expressed happy and sad at the same sentence.
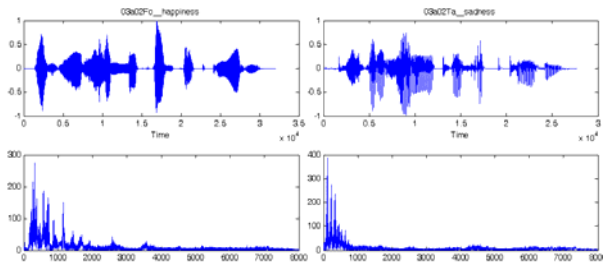

Fig. 7 The signal analysis against 2 emotions

4 The Algorithm about Speech Signals for Personalized Emotion Recognition

Proposed personalized emotion recognition algorithm using PLP analysis in this paper is divided into two parts. The first part is to extract the personalized emotion template. This template is made by PLP analysis template and the Min-Max value. The second part is to judge final emotion using the energy by PLP analysis of speech signal and Min-Max values.

The algorithm which is using energy of PLP spectrum at the speech signal and Min-Max values is shown in figure 7.

That is derived the result from PLP spectrum energy of emotion recognition and Min-Max values weight of emotion recognition. This study is optimized the weight method by Genetic Algorithm (GA). A figure 8 shows a flowchart to obtain weights.
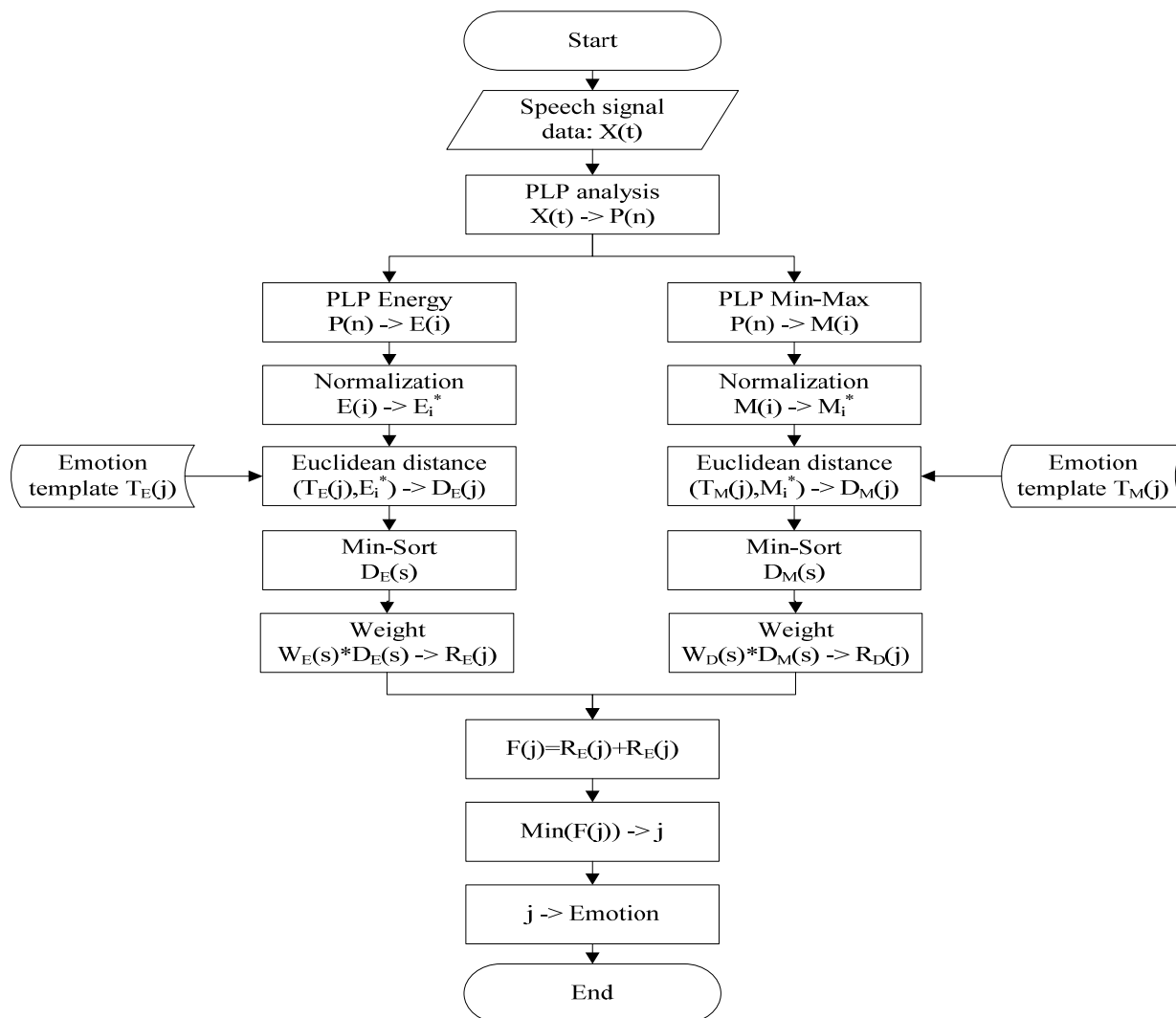

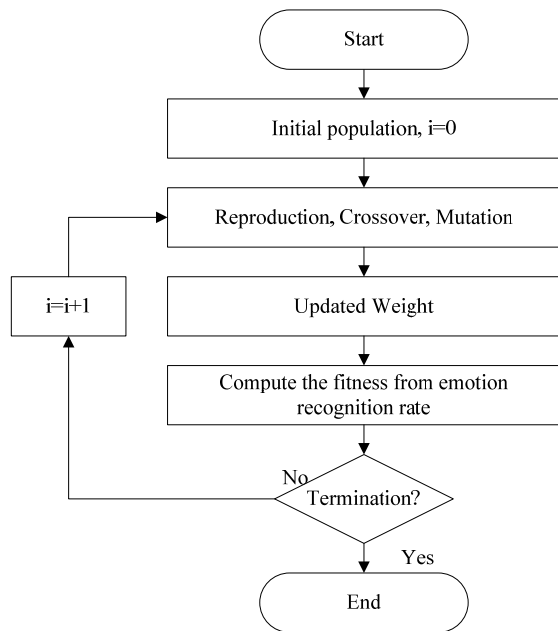Fig. 6 The Algorithm of Personalized Emotion Recognition

Fig. 8 The flowchart of GA to obtain weights

5 Result and Conclusion

The data of this experimentation is consist of 500 speech signal which are happiness, anger, anxiety, sadness, bored, disgust and neutral by five men and five women. Figure 9 shows the result which is emotional template using the algorithm in this chapter.
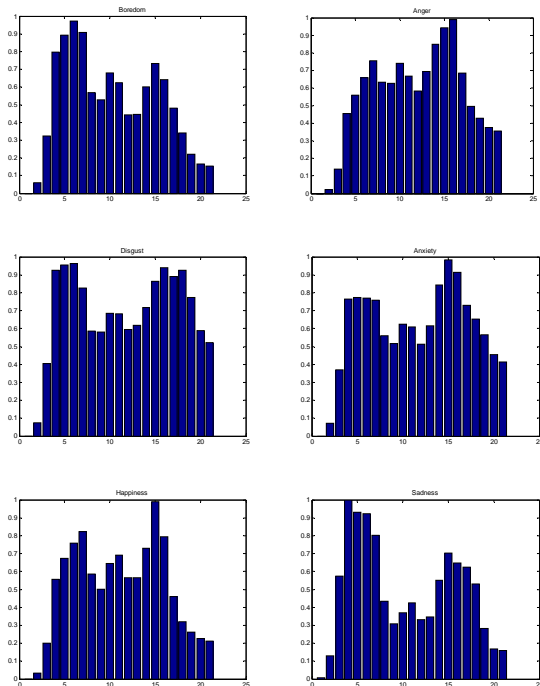

Fig. 9 PLP energy template of speech signal

This figure 9 shows personalized emotion template by PLP energy. This experimentation is applied to

make a personal template and emotion recognition as table 1. Our experimental results showed that the identification rate was over 77.1%.

Table 1 Result of Personalized Emotion Recognition of Speech Signal

Emotion	Anger	Boring	Disgust	Anxiety	Happiness	Sadness	Neutral
Anger	117	0	2	0	6	0	0
Boring	0	52	1	2	0	5	17
Disgust	3	3	32	5	2	1	4
Anxiety	3	2	4	49	2	0	2
Happiness	4	0	1	7	58	0	3
Sadness	0	2	0	3	0	52	2
Neutral	0	22	6	3	1	4	51
Total	127	81	46	69	69	62	79
Cognition Rate (%)	92.13	64.2	69.57	71.01	84.06	83.87	64.56

References
[1] Kwang-Dong Jang, Nam Kim, Oh-Wook Kwon (2006), Speech Emotion Recognition on a Simulated Intelligent Robot(in Korean), Voice 56: 173-183

[2] Chang-Hyun Park and Kwee-Bo Sim (2006), Reinforcement Learning Method Based Interactive Feature Selection(IFS) Method for Emotion Recognition(in Korean), Journal of Telecommunications and Information 12: 666-670

[3] J. Nicholson, K. Takahashi and R. Nakatsu (2000), Emotion Recognition in Speech using Neural Networks, Neural Comput & Applic 9: 290-296

[4] Aishah Abdul Razak, Ryoichi Komiya, Mohamad Izani Zainal Abidin (2005), Comparison Between Fuzzy and NN Method for Speech Emotion Recognition, Proceedings of the Third International Conference on Information Technology and Applications [5] Myun-Gu Kang, Jeong-Tae Seo, Won-Gu Kim (2004), Emotion Recognition Based on BMM by Speech Signal (in Korean), Journal of The Acoustical Society of Korea, 23: 235-241

[6] Yoon-Ho Cho, Gyu-Sik Park (2006), A Study on Robust Speech Emotion Feature Extraction Under the Mobile Communication Environment (in Korean), Journal of The Acoustical Society of Korea 25: 269-276

[7] Vladimir Hozjan and Zdravko (2003), Context-Independent Multilingual Emotion Recognition from Speech Signals, International journal of speech technology 6: 311-320