

## 음성신호기반의 감정분석을 위한 특징벡터 선택

최하나\*, 변성우\*\*, 이석필\*

상명대학교 미디어소프트웨어학과\*, 상명대학교 컴퓨터학과\*\*

### Discriminative Feature Vector Selection for Emotion Classification Based on Speech.

Ha-Na Choi\*, Sung-Woo Byun\*\*, Seok-Pil Lee\*  
SangMyung University\*, SangMyung University\*\*

**Abstract** - 최근 컴퓨터 기술이 발전하고, 컴퓨터의 형태가 다양해지면서 여러 wearable device들이 생겨났다. 이에 따라 휴먼 인터페이스 기술에서 사람의 감정정보가 중요해졌고, 감정인식에 대한 연구들이 많이 진행 되어 왔다. 본 논문에서는 감정분석에 적합한 특징벡터를 제시하고자 한다. 이를 위해 사람의 감정을 보통, 기쁨, 슬픔, 화남 4가지로 분류하고 방송매체를 통하여 잡음 없이 녹음하였다. 특징벡터는 MFCC, LPC, LPCC 3가지를 추출하였고 Bhattacharyya거리 측정을 통하여 분리도를 비교하였다.

#### 1. 서 론

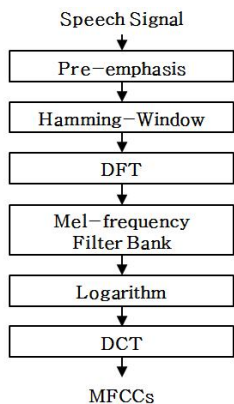
최근 컴퓨터 기술이 발전하고, 컴퓨터의 형태가 다양해지면서 여러 wearable device들이 생겨났다. 이에 따라 컴퓨터가 사람의 감정과 상황을 인식하고 그에 따른 인터랙션 작용을 하는 것이 중요해졌다. 감정을 인식하는 기술은 얼굴표정을 이용한 영상 인식기반의 기술과 EEG, PPG와 같은 생체 신호기반의 감정인식 기술, 음성신호기반의 감정인식 기술이 있고, 이런 기술들에 대한 많은 연구가 진행되어 왔다[1][2]. 특히, 음성신호는 사람의 다양한 감정정보를 가지고 있기 때문에 감정분석 연구 분야에서 얼굴인식 기반의 기술과 더불어 활발하게 연구되었다. 이러한 감정인식 기술에서 적절한 특징벡터를 선택하는 것은 시스템의 정확도를 높이는 이유로 중요하다.

본 논문에서는 음성 신호기반의 감정을 분석하기 위한 가장 적합한 특징벡터를 선택하고자 한다. 이를 위하여 사람의 감정을 보통, 기쁨, 슬픔, 화남 4가지로 분류하고 방송매체를 통하여 잡음 없이 녹음하였고, 수집한 감정데이터들을 MFCC, LPC, LPCC 3가지 특징벡터를 사용하여 분석하였다.

#### 2. 특징벡터

본 논문에서는 음성인식에 널리 쓰이고 있는 특징벡터인 Mel-Frequency cepstrum Coefficient(MFCC), Linear Predictive Coding(LPC), Linear Predictive Cepstral Coding(LPCC)를 이용하여 각 감정들간의 분리도를 비교하였다.

MFCC는 frame내의 음성 신호에 대하여 계산한 power spectrum을 청각기의 주파수 반응도를 모사한 Mel-scale 주파수 도메인에서 discrete cosine transform(DCT)를 취한 값이다. 일반적으로 이를 추출하기 위한 자세한 과정은 아래 순서도와 같다.



<그림 1> MFCC 추출 순서도

LPC는 선형 결합에 의해 과거의 신호에서 현재의 신호[n]을 예측하는 방법으로 전극(All-pole) 모델을 사용하여 식(1)과 같이 차분 방정식의 형태로 나타낼 수 있다. 여기서  $S_n$ 은 입력신호,  $\tilde{S}_n$ 은 예측신호,  $a_i$ 는 선형예측계수이며, p는 예측계수의 차수 이다. 현재신호와 예측된 신호의 예측오차는 식(2)와 같다.

$$\tilde{S} = -(a_1 S_{n-1} + a_2 S_{n-2} + \dots + a_p S_{n-p}) \quad (1)$$

$$e_n = S_n - \tilde{S}_n \quad (2)$$

식(3)은 예측오차  $e_n$ 에 대한 mean square error(MSE) J이며 J를 최소화 하는 선형예측계수를 찾기 위하여 식(3)을  $a_i$ 에 대하여 편미분하여 0이 되는 p개의 선형 연립방정식을 얻을 수 있다. 선형예측계수는 자기상관 행렬의 역행렬을 이용하여 구할 수 있다.

$$J = E[e^2(n)] = E[S_n - \tilde{S}_n^2] \quad (3)$$

LPCC는 C(z)의 inverse z-transform으로 정의되고 식은 다음과 같다.

$$C(z) = \sum_n c(n)z^{-n} \quad (4)$$

전극(All-pole)  $z = z_i$ 가 unit cycle 안에 있고, gain값을 1로 주면 LPCC ( $c_{ip}(n)$ )는 다음과 같이 정의 된다.

$$C_{ip}(n) = \frac{1}{n} \sum_{i=1}^p z_i^n \quad n > 0$$

$$= 0 \quad n < 0$$

LPCC는 recursive에 의해 예측계수 값으로부터 구한다.

$$c_1 = a_1 \quad (6)$$

$$c_n = \sum_{k=1}^{n-1} (1 - \frac{k}{n}) a_k c_{n-k} + a_n \quad 1 < n < p$$

$$c_n = \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k} \quad n > p$$

#### 3. 실험

본 논문에서 쓰이는 MFCC, LPC, LPCC 3가지의 특징벡터들의 분리도를 비교하기 위하여 Bhattacharyya거리측정[3][4]을 이용하였다.

Bhattacharyya거리 측정은 오류율을 측정하여 거리를 계산하는 방법으로 각 클래스의 분포가 가우시안 형태를 가질 때 가장 좋은 평가기준이 된다. Bhattacharyya거리의 식은 다음과 같다. 식에서  $M_1, M_2$ : 클래스 1,2의 평균,  $\sum_1, \sum_2$ : 클래스 1,2의 공분산이다.

$$\mu(1/2) = \frac{1}{8} (M_2 - M_1)^T \left\{ \frac{\sum_1 + \sum_2}{2} \right\}^{-1} (M_1 - M_2) + \frac{1}{2} \ln \frac{|\sum_1 + \sum_2|}{\sqrt{|\sum_1| |\sum_2|}} \quad (7)$$

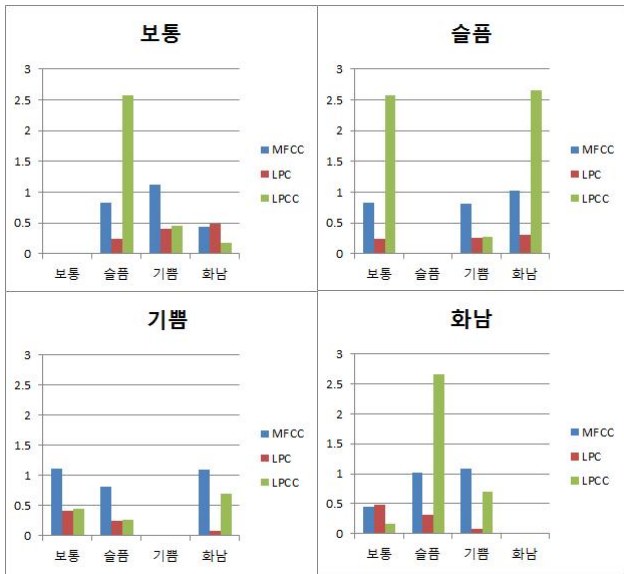
Bhattacharyya거리가 가장 큰 값이 나온 특징벡터가 클래스 간의 거리가 가장 멀리 떨어져있다는 의미로, 감정분석에 가장 적합하다.

### 3.1 실험데이터

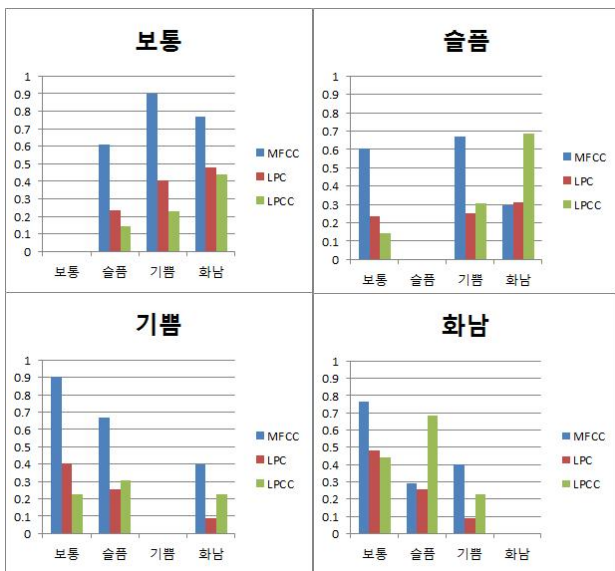
본 연구를 위한 실험 데이터는 사람에 대한 음성 감정데이터로 감정은 보통, 슬픔, 기쁨, 화남 4가지로 분류하였다. 정확한 데이터를 위하여 방송매체를 통해서 데이터를 취득하였고, 각각의 감정에 대해 3~5개씩, 총 100명의 데이터를 수집하였다. 데이터 취득 대상은 여자/남자로 구분된다. 데이터를 취득할 때 데이터들은 음성만으로 감정구분이 명확한 데이터, 배경음이 없는 데이터들로 선정하였다. 각 음성데이터들은 16000Hz로 샘플링 하였다.

### 3.2 실험결과

실험데이터들의 음성구간을 구하는 전 처리 과정을 거친 후 실험데이터들의 음성구간으로부터 특징벡터 MFCC, LPC, LPCC를 추출하였다. MFCC, LPC와 LPCC의 차수는 기존 연구들과 같이 10차로 정하였다. 이렇게 추출된 특징벡터들의 분리도를 Bhattacharyya거리를 구하여 비교하였다.



〈그림 2〉 남자음성 분리도



〈그림 3〉 여자음성 분리도

그림 2는 남자의 음성신호에서 MFCC, LPC, LPCC 3가지의 특징벡터를 추출한 후 Bhattacharyya거리를 구하여 분리도를 비교한 결과이다. 남성음성의 경우 슬픔감정과 보통감정, 슬픔감정과 화남감정사이에 분리도가 LPCC 특징벡터를 이용했을 때 약 2.5로 가장 크게 나왔다. 보통감정과 화남감정사이에 분리도는 3가지 특징벡터 모두 0.5이하로 나타났고, 기쁨 감정은 3가지 감정과의 분리도가 모두 MFCC 특징벡터를 이용했을 때가 약 1로 다른 특징벡터를 이용했을 때 보다 크게 나타났다.

그림 3은 여자의 음성신호에서 MFCC, LPC, LPCC 3가지의 특징벡터를 추출한 후 Bhattacharyya거리를 구하여 분리도를 비교한 결과이다. 여성의 보통 감정은 MFCC 특징벡터를 이용했을 때 다른 감정들과의 분리도가 가장 크고, 여성의 슬픔감정 또한 MFCC 특징벡터를 이용했을 때 분리도가 크게 나왔다. 반면, 슬픔감정과 화남감정은 LPCC 특징벡터를 이용했을 때 분리도가 가장 크게 나왔다. 기쁨감정은 보통감정과 분리도가 가장 크고, 기쁨감정 또한 MFCC 특징벡터를 이용했을 때 다른 감정들과의 분리도가 크게 나왔다. 여성의 화남감정과 기쁨 감정은 분리도가 3가지 특징벡터에서 모두 0.4이하로 나타났다.

〈표 1〉 특징벡터에 따른 Bhattacharyya 거리 값의 평균

	MFCC	LPC	LPCC
남성	0.883955	0.296184	1.138696
여성	0.60618	0.286124	0.337511

표 1은 남성과 여성의 각각의 특징벡터에 따른 Bhattacharyya 거리 값의 평균이다. 남성의 경우 평균적으로 LPCC 특징벡터를 이용했을 때 분리도가 가장 크며, 여성의 경우에는 평균적으로 MFCC를 사용했을 때 분리도가 가장 크다. 그리고 두 경우 모두 LPC를 특징벡터로 이용하였을 때 분리도가 가장 낮게 나타났다. 여성의 MFCC평균값과 남성의 LPCC평균값을 비교해 보았을 때 여성보다 남성의 음성이 분리도가 크게 나왔다.

## 4. 결 론

본 논문에서는 음성신호 기반의 감정분석을 위한 적합한 특징벡터를 선택하기 위해 실험 데이터에서 각각 MFCC, LPC, LPCC 특징벡터를 추출하여 비교, 분석 하였다. 이를 위해서 사람 25명에 대하여 보통, 슬픔, 화남, 기쁨 4가지 감정데이터를 취득하고 데이터 베이스를 구성하였다. 실험데이터에서 특징벡터를 추출한 결과를 이용하여 Bhattacharyya 거리를 구하여 각각 특징벡터에 따른 분리도를 비교하였다. 결과적으로 음성의 감정분석을 위한 특징벡터로는 남성과 여성에게 적합한 특징벡터가 각각 다르며 남성은 LPCC, 여성은 MFCC 특징벡터가 적합한 것으로 나타났다. 그리고 여성보다 남성의 음성이 분리도가 크게 나왔다.

향후, 본 논문의 실험 결과를 바탕으로 음성신호 기반의 감정분석시에 감정별로 분리도가 높은 적합한 특징벡터를 사용하여 더 효과적인 감정 분석 결과를 기대 할 수 있다.

## [참 고 문 헌]

- [1] Robert Horlings, Dragos Dacu, Leon J. M. Rothkrantz, "Emotion recognition using brain activity", CompSysTech International Conference on Computer Systems and Technologies, Article No. 6, Sept 2008.
- [2] D. O. Bos. (2006). EEG-based emotion recognition: The influence of visual and auditory stimuli. [Online]. Available: [http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude\\_Bos-Danny.pdf](http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude_Bos-Danny.pdf).
- [3] Seok-Pil Lee, Sang-Hui Park, Jeong-Seop Kim, Ig-Jae Kim "EMG pattern recognition based on evidence accumulation for prosthesis control", Proc Ann Intl Conf IEEE Eng Med Biol 4, pp 1481 - 1483, 1996.
- [4] Sang-Yeob Oh, "Improving Phoneme Recognition based on Gaussian Model using Bhattacharyya Distance Measurement Method", Journal of Korea Multimedia Society Vol 14. No.1. January 2011 pp 85-93
- [5] Jong-Yeol Yang and Hong Kook Kim "Performance Comparison of Emotion Speech Classification Methods Using MFCCs" 대한전자공학회 2008년 정기총회 및 추계종합학술대회, 2008.11, 633-634