

감정 적응을 이용한 감정 화자 인식

Emotional Speaker Recognition using Emotional Adaptation

김 원 구*
(Weon-Goo Kim)

Abstract - Speech with various emotions degrades the performance of the speaker recognition system. In this paper, a speaker recognition method using emotional adaptation has been proposed to improve the performance of speaker recognition system using affective speech. For emotional adaptation, emotional speaker model was generated from speaker model without emotion using a small number of training affective speech and speaker adaptation method. Since it is not easy to obtain a sufficient affective speech for training from a speaker, it is very practical to use a small number of affective speeches in a real situation. The proposed method was evaluated using a Korean database containing four emotions. Experimental results show that the proposed method has better performance than conventional methods in speaker verification and speaker recognition.

Key Words : Speaker recognition, Emotional speech, GMM, Emotional adaptation

1. 서 론

화자 인식 시스템의 성능은 배경 잡음, 채널 영향, 화자의 건강 상태 등 다양한 요인의 영향을 받는다. 배경 잡음과 채널 영향과 같은 외부적인 요인이나 화자의 건강 상태와 같은 내부적인 요인들은 화자 목소리에 부가적인 변화를 발생시켜 화자 인식 시스템에 부정적인 영향을 끼친다. 여기에 인간 감정은 화자 목소리에 변화를 발생시키는 또 하나의 내부적인 요인이다. 이러한 요인에 따른 화자 음성 변화는 화자 인식 시스템의 성능을 크게 저하시킨다[1-3].

감정에 따른 영향을 감소시키기 위한 방법에 대한 연구는 세 가지 영역으로 나뉘어 진행되어 왔다. 첫 번째는 감정 변화에 강인한 파라미터를 찾거나[4] 모델 학습 이전에 감정이 포함되지 않은 특징 벡터에 감정 정보를 추가하여 감정에 따른 변화를 보상하는 특징 벡터 보상 방법이다[5]. 두 번째는 화자의 감정 변화에 따른 모델 확률의 변화와 이동을 제거하는 모델 확률 보상 방법이다. 이러한 모델 확률 보상 방법의 한 예가 E-Norm이다[6]. 세 번째는 모델 영역에서 보상하는 방법으로, 감정이 포함된 화자의 음성을 사용하여 화자 인식 모델을 감정의 변화에 강인하도록 변화시키는 방법이다. 이러한 방법으로 구조적인 학습 방법[1,7], 감정 추가 모델링 방법[8], 피치와 같은 운율 정보를 사용하여 감정 음성을 분류하여 모델링하는 방법[9,10]과 감정 모델 변환 방법[11] 등이 있다.

감정 음성을 사용하여 화자 인식 모델을 감정 변화에 강인하도록 변화시키는 기존 방법들은 화자 인식 시스템의 성능을 향상

시킬 수 있었으나 다음과 같은 문제를 가지고 있다. K.R. Scherer[1]이 제안한 구조적 학습 방법은 성능 향상을 위하여 감정이 포함되지 않은 평상 음성을 사용하여 학습한 모델과 감정 음성을 사용하여 학습한 모델을 동시에 사용하였다. 이 방법은 성능이 우수한 화자 인식 모델을 만들기 위하여 많은 양의 감정 학습 데이터가 필요하다. 그러나 실제 상황에서는 많은 양의 감정 음성을 학습에 이용하는 것은 어려운 일이다. 이러한 문제점을 보완하기 위하여 W. Tian[8]이 제안한 감정 추가 모델링 방법은 운율 정보인 피치를 사용하여 14개의 감정이 포함된 학습 데이터를 세 개의 그룹으로 분류하고 평상 음성에 감정 음성을 추가하여 감정 모델을 생성하는 방법을 제안하였다. 이 방법은 3개의 화자 모델을 사용하여 성능 향상이 되었으나 여전히 많은 종류의 감정 음성이 필요하다. L. Dongdong[9]은 피치를 사용하여 평상 및 감정 음성을 구간별로 구분하여 모델링하는 방법을 사용하였다. 이 방법은 기존 방법에 비하여 적은 수의 감정 음성을 사용하였으나 학습 과정이 복잡하고 성별 인식을 위하여 추가적으로 큰 규모의 성별 인식 모델을 사용하는 단점이 있다. L. Chen[11]은 학습 과정에서 신경 회로망 등을 사용하여 평상 모델을 감정 모델로 변환하는 변환 규칙을 개발하였다. 이러한 변환 규칙은 인식단계에서 화자의 평상 모델로부터 감정 모델을 생성하는데 사용되었다. 이 방법은 제한된 학습 데이터와 변환 규칙의 정확성 부족으로 큰 성능 향상을 이루지는 못하였다.

본 논문에서는 적은 수의 감정 음성을 사용하여 화자 인식 시스템을 감정 변화에 강인하게 만들기 위하여 화자 적응 기법을 이용한 감정 모델 생성 방법을 제안하였다. 기존 방법은 많은 개수의 감정 학습 데이터가 필요하였으나 본 연구에서는 감정 적응을 위하여 적은 개수의 학습 감정 음성과 화자 적응 방식을 사용하여 감정이 없는 화자 모델로부터 감정 화자 모델이 생성되었다. 학습을 위하여 충분한 감정 음성을 얻는 것은 쉽지 않기 때

* Corresponding Author : Dept. of Electrical Engineering,
Kunsan National University, Korea.

E-mail: wgkim@kunsan.ac.kr

Received : June 1, 2017; Accepted : June 16, 2017

문에 적은 개수의 감정 음성을 사용하는 것은 실제 상황에서 매우 실용적이다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 감정 인식 시스템에 관하여 설명하고 3장에서는 본 논문에서 제안된 화자 인식 시스템의 감정 적응에 관하여 설명하였다. 4장에서는 감정 데이터를 사용한 화자 인식 실험을 수행하고 기존 시스템과 성능을 비교하였고 5장에서 결론을 맺었다.

2. 감정 화자 인식 시스템

감정 모델을 사용하여 감정 변화에 강인한 화자 인식 시스템을 개발하는 연구는 다양하게 진행되었다[1,7-10]. 이러한 방법들은 학습과정에서 감정이 포함된 화자의 음성을 사용하여 화자 인식 모델을 감정의 변화에 강인하도록 변화시키는 것이다. 본 논문에서는 제안된 방법과 성능 비교를 위하여 감정별 모델 학습, 구조적 학습 방법[1,7]과 감정 추가 학습 방법[8]을 구현하였다.

일반적으로 화자 인식 모델은 평상 음성을 사용하므로 화자마다 한 개의 모델을 갖는다, 그러나 이러한 시스템의 성능은 인식 데이터에 감정이 포함되면 급격히 저하된다. 이를 해결하는 기본적인 방법은 화자 감정별로 학습 데이터를 구축하여 감정별로 화자 인식 모델을 생성하는 것이다. 따라서 각 화자는 감정의 개수만큼 다수의 모델을 갖는다. 이러한 방법은 화자의 감정 데이터를 이용하므로 감정 변화에 강인한 시스템이지만, 실제 상황에서는 감정 학습 데이터를 충분히 확보하는 것이 쉽지 않은 문제가 있다.

구조적 학습 방법[1]은 성능 향상을 위하여 평상 음성을 사용하여 학습한 모델과 다양한 감정이 포함된 음성을 사용하여 하나의 모델을 학습하는 구조이다. 실제로 음성에 포함된 감정을 구분하는 것은 주관적이고 모호한 경우가 많다. 따라서 평상 음성과 감정 음성으로 학습 데이터를 구분하여 다양한 감정의 음성을 하나로 학습하는 것이 유리한 측면이 있다. 또한 학습을 위한 감정 데이터가 적은 경우에도 개별 모델을 만드는 경우보다 충분한 데이터로 학습이 가능한 장점이 있다. 그럼에도 불구하고 많은 양의 감정 음성을 학습에 이용하는 것은 쉽지 않고, 감정 데이터가 많아지면 다양한 감정을 모델링하기 어려워지는 문제가 있다.

감정 추가 학습[8] 방법은 피치를 사용하여 14개의 감정이 포함된 학습 데이터를 세 개의 그룹으로 분류하고 평상 음성을 추가하여 세 개의 감정 모델을 학습한다. 이러한 방법은 하나의 모델에 평상과 감정이 같이 학습되므로 학습 데이터가 부족하여 발생하는 부작용이 적어진다. 그러나 성능 향상을 위해서는 여전히 많은 감정 데이터가 필요하다. 본 연구에서는 평상이외에 3개의 감정이 포함된 데이터를 사용하므로 피치를 사용한 감정 분류는 적용하지 않고 세 개의 감정 추가 모델을 생성하여 사용하였다.

3. 화자 인식 시스템의 감정 적응

본 연구에서는 감정 음성을 사용한 화자 인식 시스템의 성능

향상을 위하여 감정 적응을 사용한 화자 인식 방법이 제안되었다. 학습 과정에서 적은 개수의 학습 감정 음성과 화자 적응 방식을 사용하여 감정이 없는 화자 모델로부터 감정 화자 모델이 생성되었다. 이러한 구조를 갖는 감정 화자 인식 시스템의 학습 과정은 그림 1과 같다.

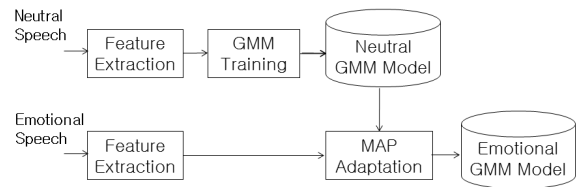


그림 1 제안된 감정 화자 인식 시스템의 학습 과정

Fig. 1 Training procedure of proposed emotional speaker recognition system

그림 1에서 각 화자의 화자 인식 모델을 생성하기 위하여 평상 음성(neutral speech)과 감정 음성(emotional speech)이 각각 사용된다. 또한 화자 인식 모델에 무감독 학습(unsupervised learning) 방법인 가우시안 혼합 모델(GMM)을 사용하였다. 우선 각 화자의 학습 데이터 중에서 감정이 없는 평상 음성을 사용하여 특징 벡터를 구한다. 특징 벡터는 평상 GMM 모델의 학습 과정에 사용되고, 이렇게 생성된 화자별 화자 인식 모델은 저장된다. 그 후 감정 음성을 사용하여 특징 벡터를 구한 후 평상 GMM 모델에 감정 적응(MAP adaptation)을 수행하여 감정 GMM 모델(emotional GMM model)을 생성한다.

기존 방법들은 감정 모델을 생성하기 위하여 감정 음성만을 사용하여 모델을 만들거나 비슷한 특성을 보이는 평상 음성과 감정 음성들을 분류하고 학습 과정을 통하여 감정 모델을 생성하였다. 본 논문에서 제안된 방법은 평상 음성을 사용한 화자 모델에 적은 개수의 감정 음성을 사용하여 화자 적응을 통하여 감정 모델을 생성하는 것이다.

감정 모델을 생성하기 위한 감정 적응을 위하여 화자 적응에 널리 사용되는 베이저안 적응(Bayesian adaptation) 또는 사후 최대 값(maximum a posteriori :MAP) 추정 방법을 사용하였다 [12]. 이러한 적응을 통해 평상 모델은 학습 감정 음성을 사용하여 감정 모델로 변환된다.

감정 적응을 위한 세부 과정은 다음과 같다. 첫 번째로 평상 음성을 사용하여 GMM 기반의 평상 GMM 모델 λ 를 생성한다. M 개의 가우시안 분포를 사용한 혼합 분포는 다음과 같이 정의된다.

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^M w_i p_i(\mathbf{x}) \tag{1}$$

가우시안 혼합 분포는 평균 벡터 μ_i , 공분산 행렬 Σ_i 과 가중 w_i 의 세 가지 파라미터가 필요하다. 이러한 파라미터로 GMM 모델을 나타내면 다음과 같다.

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, i = 1, \dots, M \tag{2}$$

두 번째로 감정 GMM 모델 $\hat{\lambda}$ 를 생성하기 위하여 감정 학습 음성을 사용하고 MAP 추정 방법을 적용하여 평상 GMM 모델 λ 의 평균, 공분산 행렬, 가중을 변환한다. 이러한 과정은 다음과 같다.

우선 평상 GMM 모델 λ 를 사용하여 감정 적응을 위한 감정 학습 음성의 특징 벡터 $X = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ 에 대한 확률 $P(i|\mathbf{x}_t)$ 와 통계 값을 구한다.

$$P(i|\mathbf{x}_t) = \frac{w_i p_i(\mathbf{x}_t)}{\sum_{j=1}^M w_j p_j(\mathbf{x}_t)} \quad (3)$$

$$n_i = \sum_{t=1}^T P(i|\mathbf{x}_t) \quad (4)$$

$$E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{t=1}^T P(i|\mathbf{x}_t) \mathbf{x}_t \quad (5)$$

$$E_i(\mathbf{x}^2) = \frac{1}{n_i} \sum_{t=1}^T P(i|\mathbf{x}_t) \mathbf{x}_t^2 \quad (6)$$

위의 확률과 통계 값들을 사용하여 평상 GMM 모델 λ 의 i 번째 혼합 분포 파라미터들을 다음과 같이 적응시켜 감정 GMM 모델 $\hat{\lambda} = \{\hat{w}_i, \hat{\mu}_i, \hat{\Sigma}_i\}$ 를 생성한다.

$$\hat{w}_i = [\alpha_i n_i / T + (1 - \alpha_i) w_i] \gamma \quad (7)$$

$$\hat{\mu}_i = \alpha_i E_i(\mathbf{x}) + (1 - \alpha_i) \mu_i \quad (8)$$

$$\hat{\Sigma}_i^2 = \alpha_i E_i(\mathbf{x}^2) + (1 - \alpha_i)(\sigma_i^2 + \mu_i^2) - \hat{\mu}_i^2 \quad (9)$$

여기서 스케일 요소 γ 는 혼합 가중의 합을 1로 만들기 위한 것이고 적응 계수 α_i 는 다음과 같이 정의되며 r 은 실험적으로 정하는 고정된 상수이다.

$$\alpha_i = \frac{n_i}{n_i + r} \quad (10)$$

이렇게 감정 적응된 화자 인식 모델들은 감정 개수만큼 생성되고 평상 음성으로 생성된 모델과 함께 화자마다 다수의 모델을 갖게 된다.

인식 단계에서는 입력 음성에 대한 유사도를 화자마다 s 개의 평상과 감정 모델들에 대하여 식 (1)을 사용하여 다음과 같이 구한다.

$$\log p(X|\lambda) = \max_{1 \leq k \leq S} \sum_{t=1}^T \log p(\mathbf{x}_t | \lambda_k) \quad (11)$$

4. 실험 및 결과

4.1 데이터베이스

화자 인식 시스템의 성능을 평가를 위하여 평상(neutral), 기쁨

(happy), 슬픔(sad)과 화남(angry)의 4가지 감정을 포함한 데이터베이스를 사용하였다. 녹음에 참여한 화자는 평소에 감정 표현 훈련이 된 남/녀 각 15명의 아마추어 연극단원으로, 녹음은 조용한 사무실 환경에서 이루어졌다. 각 화자는 45개의 서로 다른 문장을 4가지 감정으로 1회씩 녹음하였다. 사용된 문장은 0.5~3초 정도 길이의 문장으로 녹음 중에 감정 표현이 미흡하다고 판단되면 다시 녹음하였다. 본 연구에 사용된 데이터베이스는 크기는 총 5400문장(30명×45문장×4감정×1회)이다.

4.2 특징 파라미터

화자 인식을 위하여 감정이 포함된 음성 신호로부터 MFCC 파라미터를 추출하여 사용하였다. MFCC 파라미터는 다음과 같은 과정을 통하여 추출되었다. 음성 신호는 전처리 과정을 통하여 16kHz, 16비트로 샘플링되고, 고주파 성분이 보장되었다. 이렇게 샘플링된 신호는 에너지 파라미터를 사용하는 음성구간 검출 과정을 통해 묵음 구간이 제거되었다. 검출된 음성 신호로부터 20ms(320샘플)의 길이를 갖는 해밍 창을 사용하고 10ms씩 이동하면서 12차의 MFCC 파라미터가 생성되었다.

4.3 화자 인식 시스템의 구성

본 연구에서는 감정 적응을 이용한 감정 화자 인식 시스템의 성능을 평가하기 위하여 GMM 기반의 문장 독립 화자 인식 시스템을 구현하였다. 학습 과정에서는 남성 15명과 여성 15명이 총 45개의 문장 중 30개의 문장을 GMM 모델 학습에 사용하였다. 인식 과정에서는 학습에 사용되지 않은 나머지 15개 문장을 사용하였다. 학습 과정에서 평상 문장은 화자마다 30개를 사용하였고 감정 음성은 감정마다 1~30개까지 변경하면서 사용하였다. 화자 인식을 위한 화자 GMM은 64개의 혼합 분포를 갖도록 구성하였다.

본 연구에서 제안된 방법과 성능 비교를 위하여 다음과 같이 시스템을 구현하였다.

- 1) 기존 시스템 : 학습에 감정 음성을 사용하지 않음
- 2) 구조적 학습[1] : 학습 과정에서 평상 음성과 감정 음성 모델을 각각 생성
- 3) 감정 추가 학습[8] : 평상 음성에 감정 음성을 추가하여 감정 모델을 학습
- 4) 개별 모델 학습 : 학습 과정에서 동일 감정 음성으로 개별 감정 모델을 생성
- 5) 평상 모델로부터 감정 적응(제안된 방법) : 평상 음성을 사용한 GMM에 감정 음성을 사용하여 감정 적응

4.4 실험 결과

첫 번째 실험에서는 감정이 포함되지 않은 평상 음성을 사용하여 화자 인식 시스템을 학습한 후 4가지 감정이 포함된 인식 데이터를 사용하여 감정 변화에 따른 화자 인식 시스템의 성능 변화를 관찰하였다. 그림 2는 감정에 따른 GMM 기반의 화자 인식 시

스텝의 DET(Detection Error Tradeoff) 곡선을 나타낸다. 그림에서 알 수 있듯이 인식 데이터에 감정이 없는 경우(neutral)에는 가장 우수한 성능을 나타내었고 인식 데이터에 감정이 포함(happy, sad, angry)되면 인식 성능이 급격히 저하되는 것을 알 수 있다. 등가 오류율(Equal Error Rate: EER)도 평상 음성과 감정 음성 사이에 매우 큰 차이가 난다. 즉 감정이 포함되면 화자 인식 시스템의 성능이 크게 저하되는 것을 볼 수 있다.

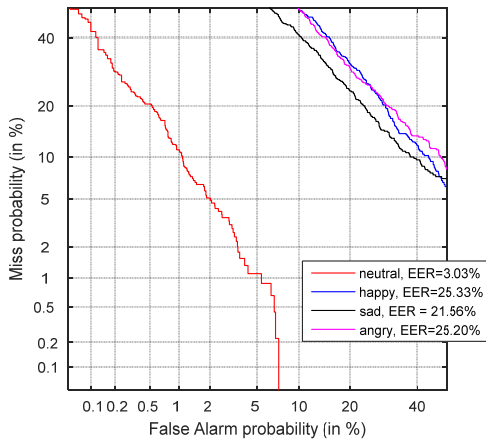


그림 2 감정 변화에 따른 화자 인식 시스템의 DET 곡선
Fig. 2 DET curves for speaker recognition system according to the emotional variation

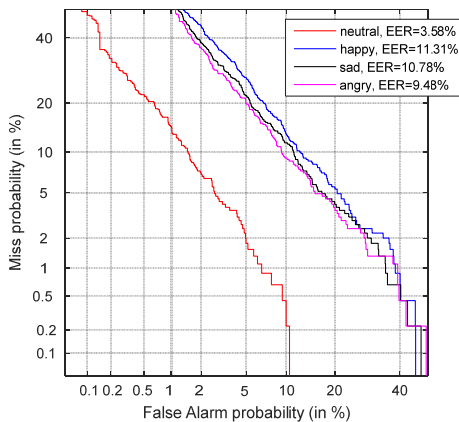


그림 3 감정 적응을 이용한 화자 인식 시스템의 DET 곡선(감정 적응에 감정마다 5문장 사용)
Fig. 3 DET curves for speaker recognition system using emotional adaptation(five utterances per emotion for emotional adaptation is used)

두 번째 실험에서는 본 연구에서 제안한 평상 음성으로 학습된 GMM 모델에 감정 음성을 사용한 감정 적응을 수행하여 감정 모델을 생성하고 화자 인식에 사용한 실험을 수행하였다. 감정 적응을 위하여 감정마다 5문장을 사용하여 평상 모델에 화자

적응을 수행하고 감정 모델을 생성하였다. 식 (10)의 적응 계수의 상수 r 은 1로 설정하였다. 그림 3은 감정 적응을 이용한 GMM 기반의 화자 인식 시스템의 DET 곡선을 나타낸다. 학습에 감정 데이터를 사용하지 않은 그림 2의 곡선에 비하여 감정마다 5개의 문장을 사용하여 감정 모델을 생성하여 사용한 결과, 감정 인식 데이터에 대한 인식 오차가 크게 감소한 것을 알 수 있다.

세 번째 실험에서는 감정 적응을 이용한 화자 인식 시스템에 대하여 감정 적응에 사용되는 감정 음성의 개수에 따른 성능 변화를 관찰하였다. 그림 4에서 감정 적응을 사용하지 않은 경우(no emotion)에 비하여 감정 적응을 위한 감정 학습 데이터의 개수를 1개부터 30개로 증가시킬수록 (1 sent.~30 sent.) 감정 적응된 화자 인식 시스템의 성능이 향상되는 것을 알 수 있다. 본 실험에 사용된 시스템은 문장 독립 화자 인식 시스템이므로 충분한 감정 적응을 위해서는 20 문장 이상이 사용되어야 하는 것으로 볼 수 있다.

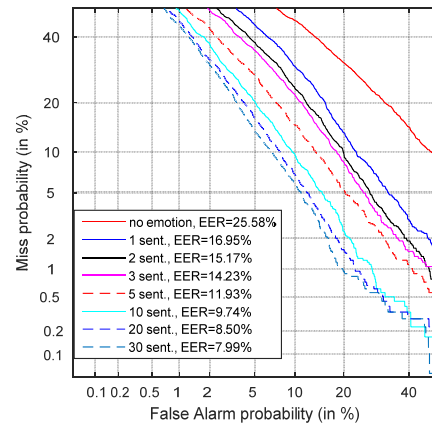


그림 4 감정 적응 데이터 개수에 따른 감정 적응을 이용한 화자 인식 시스템의 DET 곡선
Fig. 4 DET curves for speaker recognition system using emotional adaptation according to the number of adaptation data

네 번째 실험에서는 본 연구에서 제안된 감정 변화에 강인한 화자 인식 시스템의 성능을 기존 시스템과 비교하였다. 실험에 사용된 시스템은 학습에 감정 음성을 사용하지 않는 기존 시스템(traditional), 학습 과정에서 평상 음성과 감정 음성을 각각 생성하는 구조적 학습 시스템(structure training), 학습과정에서 평상 음성에 감정 음성을 추가하여 감정 모델을 생성하는 감정 추가 학습 모델 시스템(emotion-added model)과 학습 과정에서 감정별로 모델을 생성하는 개별 모델 시스템(separate model)이다. 이들을 본 연구에서 제안된 감정 음성을 사용하여 감정이 없이 학습된 모델을 감정 적응하여 감정 모델로 변환하는 방법과 성능을 비교하였다. 그림 5는 학습 과정에서 각 감정 모델마다 사용된 감정 음성의 개수가 3개(a)인 경우와 5개(b)인 경우에 각 시스템의 성능을 나타낸다. 두 경우 모두 제안된 방법이 가장 우수한 성능을 나타내었다.

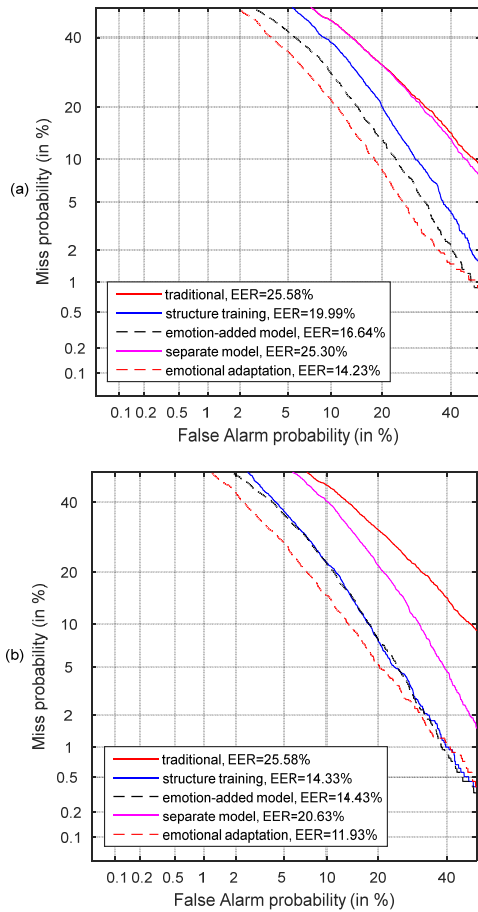


그림 5 기존, 구조적 학습, 감정 추가 학습 모델, 개별 모델과 감정 적응 화자 인식 시스템의 DET 곡선, 감정마다 사용된 감정 발성의 수: (a) 3개 (b) 5개

Fig. 5 DET curves for traditional recognition system, structure training system, emotion-added model system, separate modeling system and emotional adaptation system, number of affective utterance per emotion: (a) three (b) five

마지막 실험에서는 기존 시스템과 제안된 시스템의 화자 식별 (speaker identification) 실험을 수행하였다. 그림 6은 학습 과정에서 감정별로 사용된 감정 음성의 개수에 따른 화자 식별 시스템의 성능을 비교하였다. 학습에 사용된 감정 음성의 개수가 적을수록 성능 저하가 크게 발생하며 감정의 개수가 증가함에 따라 감정에 따른 오차가 감소한다. 특히 제안된 방법은 학습에 사용된 감정의 개수가 적을 때 기존 시스템에 비하여 우수한 성능을 보이고 있다. 실제로 화자로부터 학습을 위하여 충분한 감정 음성을 얻는 것은 쉽지 않기 때문에 적은 개수의 감정 음성을 사용하는 것은 실제 상황에서 매우 실용적이다. 학습과정에서 감정별로 독립적으로 감정 모델을 생성하여 사용하는 개별 모델 시스템(separate model)은 감정 데이터가 감정별로 15개 이상 충분히 주어지야만 우수한 성능을 나타낸다.

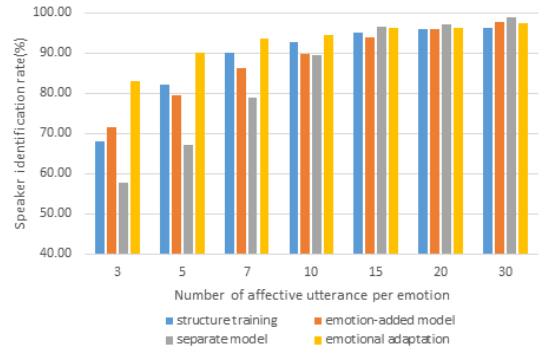


그림 6 학습에 사용된 감정별 감정 발성 개수와 학습 방법에 따른 감정 화자 식별 시스템의 성능 비교

Fig. 6 Performance comparison of speaker identification systems according to the number of affective utterance per emotion

5. 결 론

다양한 감정이 포함된 음성은 화자 인식 시스템의 성능을 저하시킨다. 본 논문에서는 감정 음성을 사용한 화자 인식 시스템의 성능 향상을 위하여 감정 적응을 사용한 화자 인식 방법이 제안되었다. 감정 적응을 위하여 적은 개수의 학습 감정 음성과 화자 적응 방식을 사용하여 감정이 없는 화자 모델로부터 감정 화자 모델이 생성되었다. 화자로부터 학습을 위하여 충분한 감정 음성을 얻는 것은 쉽지 않기 때문에 적은 개수의 감정 음성을 사용하는 것은 실제 상황에서 매우 실용적이다.

제안된 방법은 4가지 감정이 포함된 한국어 데이터베이스를 사용하여 평가되었다. 본 연구에서 제안된 감정이 없이 학습된 모델을 감정 음성을 사용하여 감정 적응하여 감정 모델로 변환하는 방법과 기존 학습 방법의 성능을 비교하였다. 실험 결과에서 제안된 방법은 학습 과정에 사용되는 감정 학습 데이터 개수가 적을수록 기존 시스템과의 성능 차이가 크게 났다.

제안된 방법은 학습에 사용된 감정 데이터 개수에 따른 화자 식별 실험에서도 적은 개수의 감정 음성을 사용하면서 기존 시스템보다 우수한 성능을 나타내었다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부) 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2012 R1A1A4A01014421).

References

[1] K. R. Scherer, T. Johnstone, G. Klasmeyer, and T. Bänziger,

“Can Automatic Speaker Verification be Improved by Training the Algorithms on Emotional Speech?”, ICSLP 2000, pp. 807-810, 2000.

- [2] I.R. Murray, J.L. Arnott, “Applying an Analysis of Acted Vocal Emotions to Improve the Simulation of Synthetic Speech”, Computer Speech and Language 22(2), pp. 107-129, 2008
- [3] B. Huanjun, X. Mingxing, and F. Z. Thomas, “Emotion Attribute Projection for Speaker Recognition on Emotional Speech”. EUROSPEECH 2007, Antwerp, pp. 758-761, 2007.
- [4] K. Alluri, V. Raju, S. Gangashetty, A. Kumar Vuppala “Analysis of Source and System features for Speaker Recognition in Emotional Conditions”, 2016 IEEE Region 10 Conference (TENCON), pp. 2847-2850, 2016
- [5] L. Dongdong, Y. Yingchun, and W. Zhaohui, “Emotion-State Conversion for Speaker Recognition”, ACII 2005, Beijing, pp. 403-410, 2005.
- [6] W. Wei, F.Z. Thomas, M. X, and B. HuanJun, “Study on Speaker Verification on Emotional Speech”, Proc. Interspeech '2006, pp. 2102-2105, 2006.
- [7] L. Macková, A. Čížmár, “Emotional Speaker Verification Based on I-vectors”, 5th IEEE International Conference on Cognitive Infocommunications, pp. 533-536, 2014.
- [8] W. Tian, Y. Yingchun, and W. Zhao-hui, “Improving Speaker Recognition by Training on Emotion-Added Models”, ACII 2005, Beijing, pp. 382-389, 2005.
- [9] L. Dongdong and Y. Yingchun, “Emotional Speech Clustering based Robust Speaker Recognition System,” the 2nd International Congress on Image and Signal Processing, pp. 1-5, Tianjin, China, Oct. 2009.
- [10] L. Dongdong and Y. Yingchun, D. Weihui, “Cost-Sensitive Learning for Emotion Robust Speaker Recognition”, The Scientific World Journal, Vol. 2014, pp. 1-9, 2014.
- [11] L. Chen, Y. Yang, “A preliminary study on GMM weight transformation for Emotional Speaker Recognition”, 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 31-36, 2013
- [12] Reynolds, D.A., Quatieri, T.F., Dunn, R.B., “Speaker Verification using Adapted Gaussian Mixture Models”, Digital Signal Processing 10, pp. 19-41, 2000.

저 자 소 개



김 원 구 (Weon-Goo Kim)

1987년 연세대학교 전자공학과 졸업, 1989년 동 대학원 전자공학과 공학석사, 1994년 동 대학원 전자공학과 공학박사, 1998-1999년 Bell Lab, Lucent Technologies(USA) 객원연구원, 2008-2009년 Griffith University 방문 교수, 1994년~현재 군산대학교 전기공학과 교수.