한국음향학회지 제22권 제7호 pp. 519~527 (2003)

화자식별을 위한 강인한 주성분 분석 가우시안 혼합 모델

RPCA-GMM for Speaker Identification

이 윤 정*, 서 창 우**, 강 상 기***, 이 기 용*
(Youn-Jeong Lee*, Chang-Woo Seo**, Sang-Ki Kang***, Ki-Yong Lee*)

*숭실대학교 정보통신전자공학부, **인스 모바일 기술연구소, ***삼성전자 정보통신총괄 연구소

(접수일자: 2003년 4월 4일; 수정일자: 2003년 8월 8일; 채택일자: 2003년 8월 16일)

음성신호는 주변 잡음과 화자의 발성 패턴 변화, 음성 검출 오류에서 생기는 이상치 (outlier)에 많은 영향을 받고 있다. 이러한 음성 신호를 이용하여 화자인식에 이용할 경우 인식률이 저하된다. 본 논문에서는 화자식별 (speaker identification)에서 학습 특징 벡터의 이상치와 고차원 문제를 해결하기 위하여 M-추정을 이용한 강인한 주성분 분석 가우시안 혼합 모델 (Robust Principal Component Analysis-Gaussian Mixture Model)방법을 제안하였다. 제안된 방법은 먼저, 특징 벡터에 이상치가 존재할 경우 M-추정에 의하여 강인한 공분산 행렬을 재추정하여 얻어진 고유벡터로부터 변환 행렬을 구하여 감소된 차원을 갖는 새로운 특징벡터를 구한다. 여기에서 얻은 선형변환된 특징벡터로부터 화자의 가우시안 혼합 모델을 구한다. 제안된 방법의 성능을 검증하기 위하여 화자식별 실험을 하였다. 실험은 전형적인 가우시안 혼합 모델 방법과 주성분 분석법, 제안된 방법을 비교 분석하였다. 이상치가 2%씩 증가할 때마다 가우시안 혼합 모델 방법과 주성분 분석법은 각각 0.65%, 0.55%씩 화자식별 성능이 저하되었지만, 제안된 방법은 0.03% 정도 감소하였으므로 이상치에 더욱 강인함을 알 수 있다.

핵심용어: 화자식별 , 주성분 분석법, 가우시안 혼합 모델, M-추정
투고분야: 음성처리 분야 (2,5)

Speech is much influenced by the existence of outliers which are introduced by such an unexpected happenings as additive background noise, change of speaker's utterance pattern and voice detection errors. These kinds of outliers may result in severe degradation of speaker recognition performance. In this paper, we proposed the GMM based on robust principal component analysis (RPCA-GMM) using M-estimation to solve the problems of both outliers and high dimensionality of training feature vectors in speaker identification. Firstly, a new feature vector with reduced dimension is obtained by robust PCA obtained from M-estimation. The robust PCA transforms the original dimensional feature vector onto the reduced dimensional linear subspace that is spanned by the leading eigenvectors of the covariance matrix of feature vector. Secondly, the GMM with diagonal covariance matrix is obtained from these transformed feature vectors. We performed speaker identification experiments to show the effectiveness of the proposed method. We compared the proposed method (RPCA-GMM) with transformed feature vectors to the PCA and the conventional GMM with diagonal matrix. Whenever the portion of outliers increases by every 2%, the proposed method maintains almost same speaker identification rate with 0.03% of little degradation, while the conventional GMM and the PCA shows much degradation of that by 0.65% and 0.55%, respectively. This means that our method is more robust to the existence of outliers.

Keywords: *Speaker identification, Principal component analysis, GMM, M-estimation*
ASK subject classification: *Speech signal processing (2,5)*

I. 서 론

사람이 말하는 음성신호에는 음운정보뿐만 아니라, 각 개개인의 독특한 생체정보를 가지고 있다. 각 개개인의 생체적 정보로서의 음성 신호를 이용하여 누구의 음성인지 알아내는 방법을 화자인식이라 한다.

화자인식은 화자식별 (speaker identification: SI)과 화자확인 (speaker verification: SV)으로 나눌 수 있다. 화자식별은 발성된 음성 신호가 등록된 화자들 중에서

†임저자: 이윤정 (yjlee@clsp.ssu.ac.kr)
156-743 서울시 동작구 상도5동 1-1
숭실대학교 정보통신전자공학과
(전화: 02-817-4591; 팩스: 02-817-4591)

어떤 화자인지 골라내는 것이다. 화자확인은 발성된 음성신호가 등록된 화자의 음성과 일치하는지를 판정하는 것으로, 발성한 화자와 등록된 화자와의 확인과정을 통하여 문턱값 (threshold)보다 유사도가 큰 경우 수락(accept)하고, 문턱값보다 유사도가 작은 경우 거절(reject)하는 것이다.

화자인식 방법을 발성방법에 따라 분류하면 문장 종속(text-dependent)형과 문장 독립 (text-independent)형이 있다. 문장 종속형 화자인식은 학습과정과 테스트를 위하여 미리 정해놓은 단어나 문장을 사용하고, 문장 독립형 화자인식은 학습과정과 테스트과정에서 발성하는 음성신호에 제한을 두지 않는다. 문장 종속 및 독립형은 녹음기 등을 이용하여 등록된 화자의 음성을 통해 등록을 시도할 때 등록된 화자로 인식되는 문제점이 발생하므로 단어나 숫자의 나열이 아닌 임의의 문장을 제시하여 검증하는 문장 제시형 (text-prompted)이 등장하였다.

화자인식에는 DTW (Dynamic Time Warping) 알고리즘, HMM (Hidden Markov Model) 방법, 그리고 GMM 방법 등이 사용된다. DTW는 입력된 음성신호가 기준 패턴과 일치하는 정도를 이용하는 방법이다. 비교적 짧은 문장을 사용하므로 구현이 쉽고 용이하지만 다른 사람이 비슷한 발성을 한 경우에 인증이 될 수 있고, 화자의 감정변화나 발성 패턴 변화, 억양 등이 변할 때 인증률이 저하되는 단점이 있다. HMM은 Markov 연쇄에 기초한 시계열 패턴의 생성과정이 각 상태에서 출력값이 결정론적으로 정해지는 것이 아니라 각 출력 값이 출력할 확률만으로 지정되는 방법이다. HMM 방법은 DTW보다 긴 문장에 사용이 가능하고 어떤 문장이나 발성이 가능하지만, 파라메타 수가 많이 필요하여 적은 데이터에 사용이 불가능하고 학습된 패턴에 종속적이라는 단점이 있다. GMM은 출력 밀도 함수가 한 개의 상태로만 이루어지는 CHMM (Continuous HMM)의 한 형태로 여러 혼합성분 들의 가우시안 확률 분포를 사용하여 화자인식에 사용한다. GMM은 문장 독립형을 주로 사용하지만 다른 문장을 발성시 인증되는 문제로 인하여 점차 문장 제시형으로 많이 사용되고 있다.

본 논문에서는 화자인식에 최근 많이 사용되고 있는 GMM 방법을 사용한다. GMM 방법은 벡터의 요소 (element) 들 사이에 상관관계가 존재하지 않는다는 가정 하에 공분산의 대각 (diagonal) 성분만을 이용하여 화자식별과 화자확인에 많이 사용되고 있지만[1], 실제로 벡터의 요소들 사이에는 상관관계가 존재하므로 화자인식의 성능 저하를 가져온다. 화자인식의 성능을 높이기 위해서는 많은 혼합성분 개수와 높은 차원의 특징 벡터가 필요하다[1,2]. 그러나 많은 혼합성분 개수와 높은 차원의 특징 벡터는 많은 음성데이터를 필요로 하며 계산과정이 복잡해지고 실시간 구현을 어렵게 한다[3].

이러한 GMM 방법의 문제점을 해결하기 위하여 주성분 분석 (Principal Component Analysis : PCA)에 기반을 둔 직교 (orthogonal) GMM 방법이 제안되었다[2]. 주성분 분석은 특징 벡터들의 고유값 (eigenvalue)과 고유벡터 (eigenvector)로부터 변환 행렬 (transform matrix)을 구하여 정보의 손실없이 요소들 사이에 서로 독립인 감소된 차원을 갖는 주성분 벡터로 축약시키는 방법이다[4]. 그러나 특징벡터의 주성분 분석은 화자자신의 불규칙한 발성 패턴이나 강세, 억양 등이 갑자기 변화할 때, 잘못된 음성 검출시 발생하는 신호와 주변 잡음같은 이상치(outlier)에 상당히 민감하므로 화자의 순수한 특징 벡터만을 추출하기 어렵다[3]. 따라서 본 논문에서는 반복적인 M-추정에 의하여 이상치의 영향을 감소시키는 변환 행렬을 구하고 강인한 주성분 분석을 통하여 선형변환된 주성분 벡터를 화자인식 모델을 위한 GMM에 사용한다.

본 논문의 구성은 다음과 같다. 화자인식을 위하여 II장에서는 일반적인 주성분 분석과 강인한 주성분 분석에 대하여 정의한다. III장에서는 강인한 주성분 분석을 이용한 GMM 방법, IV장에서는 화자식별에 대하여 기술한다. V장에서는 200명의 음성 데이터를 사용하여 전형적인 GMM 방법과 주성분 분석법 및 제안된 방법을 비교 분석하고, 마지막으로 결론을 내린다.

II. 강인한 주성분 분석

2.1. 주성분 분석을 이용한 주성분 벡터 추출

주성분 분석은 여러 개의 변수들에 대하여 얻어진 다변량 자료를 분석대상으로 하여 다차원적인 변수들을 축소, 요약하는 차원의 단순화와 함께 일반적으로 서로 상관관계가 있는 반응 변수들간의 복잡한 구조를 분석하는데 목적이 있다. 따라서 입력된 음성데이터로부터 추출된 특징 벡터들을 상관관계가 없는 새로운 좌표계로 선형변환시켜 좌표 변환에 의해 새롭게 변형된 성분을 계산한다.

즉, $\vec{X}_t = [x_1, x_2, \cdots, x_p]$, $t = 1,2,\cdots,T$ 인, p-차원 특징 벡터열 $X = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_T\}$ 가 주어졌을 때, 정보의 손실 없이 특징 벡터 $\vec{X}_t$의 요소 사이에 존재하는 상관 관계를 제거하여 독립된 요소를 갖는 새로운 k-차원 $(k \le p)$ 주성분 벡터열 $Y = \{\vec{Y}_1, \vec{Y}_2, \cdots, \vec{Y}_T\}$ 을 구하는 것이다.

여기서, $\vec{Y}_t = [y_1, y_2, \cdots, y_k]$ 이다.

주성분 벡터열 Y를 구하기 위하여 먼저, 주어진 특징 벡터의 평균 벡터 $\vec{C}$ 와 공분산 행렬 Σ_X을 먼저 구한다.

$$\vec{C} = \frac{1}{T}\sum_{t=1}^{T}\vec{X}_t \tag{1}$$

$$\Sigma_X = \frac{1}{T}\sum_{t=1}^{T}(\vec{X}_t - \vec{C})(\vec{X}_t - \vec{C})^T \tag{2}$$

여기에서, 공분산 행렬 Σ_X는 고유값과 고유벡터로 분해 된다.

$$\Sigma_X = \sum_{i=1}^{p}\lambda_i \upsilon_i \upsilon_i^T \tag{3}$$

λ_i는 Σ_X의 i번째 고유값이고, υ_i는 변환행렬 Ω의 i번째 열벡터인 고유벡터이고, 고유값 λ_i, $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p)$, $i = 1, 2, \cdots, p$이 주어졌을 때 정대응된다. Ω는 $p \times p$인 직교 행렬 $(\Omega\Omega^T = I)$을 이룬다. 이로부터 t번째 상태열의 특징 벡터 $\vec{X}_t$와 주성분 벡터 $\vec{Y}_t$의 관계는

$$\vec{Y}_t = \Omega^T \vec{X}_t \tag{4}$$

로 나타낼 수 있다. 여기에서 Ω^T는 X를 Y로 선형변환하기 위한 크기가 $p \times p$인 변환 행렬이다.

만약 $k = p$이면, 위의 식은 $\vec{X}_t$에서 $\vec{Y}_t$으로 손실없이 원 자료의 전체 공분산을 100% 표현한다. 고유값의 크기순 $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_p, k < p)$으로 나타낼 때, λ_i의 짝을 υ_i으로 나타내어 k-차원에 해당되는 Ω_s를 선택한다

$$\Sigma_Y = \sum_{i=1}^{k}\lambda_i \upsilon_i \upsilon_i^T, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{5}$$

$$\Omega_S = [\upsilon_1 \quad \upsilon_2 \quad \upsilon_3 \quad \cdots \quad \upsilon_k] \tag{6}$$

k-차원 주성분 벡터의 정보 비율(I)은 다음 식에 의해 구할 수 있다.

$$I = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{p}\lambda_i} \tag{7}$$

이 정보비율에 따라 고유값이 큰 것부터 k-차원만을 선택하여 Ω_s^T를 구하고,

$$\vec{Y}_t = \Omega_S^T \vec{X}_t \tag{8}$$

로 주성분 벡터를 얻는다.

2.2. 강인한 주성분 분석을 이용한 주성분 벡터 추출

일반적인 주성분 분석법은 정보의 손실없이 특징 벡터 $\vec{X}_t$를 독립된 요소로 구성된 새로운 차원을 갖는 주성분 벡터열을 구하는 것이다. 따라서 새로운 차원의 주성분 벡터를 이용하여 화자인식에 적용할 경우 데이터 저장 공간이 적게 필요하고 계산량이 줄어드는 이점이 있다.

그러나 화자자신의 불규칙한 발성 패턴이나 강세, 억양 등이 갑자기 변할 때 발생하는 신호와 주변 잡음, 잘못된 음성 검출시 발생하는 이상치가 특징 벡터들에 포함된 경우 일반적인 주성분 분석법은 이상치에 의해 특징 벡터들의 평균이 이동되어 분산 값이 변하고, 잘못된 변환행렬을 얻게 되므로 주성분 벡터로의 선형변환이 정확하지 않게 되어 정확한 모델링을 하기 어렵다. 따라서 특징 벡터에 이상치가 존재할 경우 일반적인 주성분 분석법의 보완이 필요하게 된다.

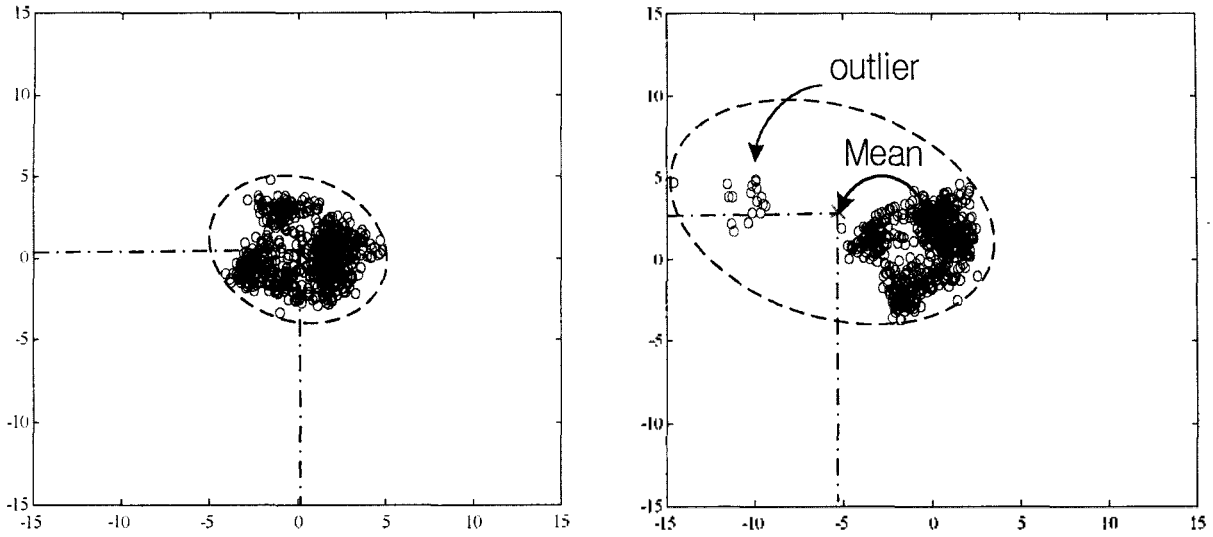
그림 1은 특징 벡터와 평균값의 분포를 나타낸 것이다. (a)는 원음성에서 특징 벡터를 추출한 경우의 분포와 평균을 나타낸 것이고, (b)는 특징벡터에 이상치가 존재하는 경우의 특징 벡터와 평균을 나타낸 것이다. (b)는 (a)에 비해 평균과 분산이 이동됨을 알 수 있다. 즉 특징 벡터에 이상치가 존재하면, 이상치가 존재하지 않는 경우와 달리 정확하지 않은 주성분 벡터가 생성된다. 주성분 분석이 제대로 되기 위해서는 정확한 평균과 분산을 구해야 하는데, 특징 벡터에 이상치가 포함될 경우 이상치의 영향을 받아 불안정한 평균과 분산이 구해진다. 따라서 본 논문에서는 이상치가 존재하여도 이상치의 영향에 좌우되지 않는 M-추정에 기반을 둔 강인한 평균과 공분산을 반복적으로 구하였다. 강인한 평균과 공분산으로부터 강인한 주성분 분석을 다음과 같이 구한다.

평균 벡터 $\vec{C}$와 분산 벡터 $\vec{V}$를 이용하여 특징 벡터와 평균 벡터 사이의 거리 d_t를 측정한다.

$$d_t = \sum_{i=1}^{p}\frac{(x_{t,i} - \vec{C}_i)^2}{\vec{V}_i} \tag{9}$$

d_t를 이용하여 경계값과 비교하여, 이상치의 영향을 감소시키기 위하여 Huber weight 함수[5]를 사용하였다.

$$\begin{cases} w_t = 1 & , \quad d_t \leq q_\delta \\ w_t = \dfrac{q_\delta \operatorname{sgn}(d_t)}{d_t} & , \quad d_t > q_\delta \end{cases} \tag{10}$$



(a) 원음성의 특징벡터들
(a) The feature vectors in the original speech

(b) Outlier가 있는 특징벡터들
(b) The feature vectors with outliers

그림 1. Outlier의 유무에 따른 특징벡터의 분포
Fig. 1. The distribution of feature vectors with outliers.

여기에서, q_s는 강인한 평균 벡터를 위한 경계값으로

$$q_s = \beta \cdot \frac{1}{T}\sum_{t=1}^{T} d_t \tag{11}$$

여기서 β는 경계값의 가중치이다.

d_t가 q_s보다 큰 경우인 이상치가 포함된 특징벡터는 영향을 줄여주고, d_t가 q_s보다 작은 경우인 이상치가 포함되지 않은 특징 벡터는 그대로 사용하였다.

위의 식을 이용하여 잡음과 같은 이상치에 강인한 평균과 공분산 행렬을 재추정할 수 있다.

$$\hat{\vec{C}} = \frac{\sum_{t=1}^{T} w_t \vec{X}_t}{\sum_{t=1}^{T} w_t} \tag{12-a}$$

$$\hat{\Sigma}_X = \frac{\sum_{t=1}^{T} w_t (\vec{X}_t - \hat{\vec{C}})(\vec{X}_t - \hat{\vec{C}})^T}{\sum_{t=1}^{T} w_t} \tag{12-b}$$

재추정 과정을 몇 번 반복하여 이상치의 영향을 감소시킨후, 식 (12-b)의 공분산 행렬로부터 이상치의 영향에도 강인한 변환행렬 $\hat{\Omega}_s'$를 구할 수 있다. t번째 상태열의 특징 벡터 $\vec{X}_t$와 주성분 벡터 $\vec{Y}_t$의 관계는 다음 식으로 표현된다.

$$\vec{Y}_t = \hat{\Omega}_s' \vec{X}_t \tag{13}$$

III. 강인한 주성분 분석 가우시안 혼합 모델

3.1. GMM 학습 과정

p-차원을 가지는 상태열 T개의 학습 벡터를 $X = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_T\}$라 두자. II에서 제안한 강인한 주성분 분석법을 이용하여 주성분 벡터의 차원을 k라 가정하자. 이 주성분 벡터를 사용한 가우시안 성분 밀도 함수는 성분의 가중치 (weight), 평균벡터 (mean vector), 분산 행렬 (variance matrix)로 나타낼 수 있다.

$$\theta = \{p_i, \vec{\mu}_i, \Sigma_i\}, \quad i = 1, \ldots, M \tag{14}$$

$k \times p$ 변환 행렬을 사용하여 식 (13)에 의해 T번째열의 k-차원 학습 주성분 벡터 $\vec{Y}_t = [y_1, y_2, \cdots, y_k]$를 구한다. 여기서 $t = 1,2,\cdots,T$ 이다. 변환된 k-차원 주성분 벡터 Y를 이용한 GMM의 유사도는

$$p(Y|\lambda) = \prod_{t=1}^{T} p(\vec{Y}_t|\theta) \tag{15}$$

로 구할 수 있다[1]. $p(\vec{Y}_t|\theta)$는 성분(mixture)의 확률 밀도값의 가중된 합이다.

$$p(\vec{Y}_t|\theta) = \sum_{i=1}^{M} p_i b_i(\vec{Y}_t) \tag{16}$$

여기서, $b_i(\vec{Y}_t)$는 k-차원 가우시안 성분 밀도 (Gaussian mixture density)이고, 각각의 성분 밀도값은 평균 벡터 $\vec{\mu}_{Yi}$ 와, 공분산 행렬 Σ_{Yi} 로 나타낸다.

$$b_i(\vec{Y}_t) = \frac{1}{(2\pi)^{k/2}|\Sigma_{Yi}|^{1/2}} \exp\left[-\frac{1}{2}\left\{(\vec{Y}_t - \vec{\mu}_{Yi})' \Sigma_{Yi}^{-1} (\vec{Y}_t - \vec{\mu}_{Yi})\right\}\right] \quad (17)$$

GMM의 확률 값을 최대로 하기 위해 ML (Maximum Likelihood) 알고리즘을 사용하여 $\frac{\partial p(Y|\theta)}{\partial p_i} = 0$, $\frac{\partial p(Y|\theta)}{\partial \mu_{Yi}} = 0$, $\frac{\partial p(Y|\theta)}{\partial \Sigma_{Yi}} = 0$ 를 만족하는 모델 파라메터 θ 를 찾는다. 그러나 ML 알고리즘을 만족하는 파라메타를 직접적으로 구할 수 없다. 그러므로 ML 파라메타 추정은 반복적으로 EM (Expectation- maximization) 알고리즘을 사용하여 얻는다[7]. EM 알고리즘은 초기 모델 θ 로부터 $p(Y|\bar{\theta}) \geq p(Y|\theta)$ 인 새로운 모델 $\bar{\theta}$ 를 추정하는 것이다. 다음 반복 과정에서 새로운 모델이 다시 초기 모델이 되고, 이러한 과정을 수렴값으로 수렴할 때까지 반복적으로 수행한다. 따라서 EM 알고리즘을 반복하는 동안 모델의 유사도 값이 단조 증가되는 가중치, 평균벡터, 분산 행렬을 다음 식으로 재추정한다.

• 성분의 가중치 (weight)

$$\bar{p}_i = \frac{1}{T}\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta) \quad (18)$$

• 평균 벡터 (mean vector)

$$\bar{\vec{u}}_{Yi} = \frac{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)\vec{Y}_t}{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)} \quad (19)$$

• 분산 행렬 (variance matrix)

$$\bar{\Sigma}_{Yi} = \frac{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)\vec{Y}_t^2}{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)} - \bar{\vec{u}}^2_{Yi} \quad (20)$$

• 사후확률 (A posterior probability)

$$p(i|\vec{Y}_t,\theta) = \frac{p_i b_i(\vec{Y}_t)}{\sum_{m=1}^{M} p_m b_m(\vec{Y}_t)} \quad (21)$$

음성의 특징 벡터열 X로부터 강인한 주성분 분석법을 사용하여 선형변환된 주성분 벡터 Y를 이용하여 다음 단계별로 GMM 학습이 이루어진다.

Step 1.

특징 벡터열 X로부터, 화자의 강인한 주성분 분석법을 이용하여 주성분 벡터 Y로 선형변환시킨다.

Step 2.

$i = 1,\cdots,M$ 일 때, 주성분의 벡터 Y의 사후확률 $p(i|\vec{Y}_t,\theta)$ 를 구한다.

Step 3.

$i = 1,\cdots,M$ 일 때, 식 (18~21)을 이용하여 Y의 파라메타 $\theta = \{p_i, \bar{\mu}_{Yi}, \Sigma_{Yi}\}$의 요소인 가중치 $\bar{p}_i$, 평균 $\bar{\vec{u}}_{Yi}$, 분산 Σ_{Yi}을 계산한다.

Step 4.

Y의 사후확률 $p(i|\vec{Y}_t,\theta)$을 다시 계산하여 GMM 유사도를 계산한다. GMM 유사도 값이 정해진 수렴값보다 크면 step 3~step 4까지 반복하고 수렴값보다 작으면 멈춘다.

그림 2는 위의 단계를 나타낸 것이다. 학습을 위한 음성 신호가 입력되면, 먼저 특징 벡터들의 평균과 분산을 구하여 이상치의 영향을 줄이기 위한 화자의 강인한 주성분 분석법을 구한다. 강인한 주성분 분석법의 변환행렬을 사용하여 특징 벡터의 차원을 줄이고, GMM 화자 학습 모델을 구한다.

IV. 화자식별

화자식별은 발성된 음성 신호가 등록되어 있는 화자들 중에서 가장 유사도가 높은 화자를 골라내는 것이다. S명의 화자로부터, 화자 각각은 강인한 주성분 분석법을 갖는 GMM의 $\theta_1,\theta_2,\cdots,\theta_s$로 나타낸다. 화자의 음성이 입력되면, 특징 벡터 $\vec{X}_t$는 학습시 저장된 각 화자의 강인한 주성분 분석법을 이용하여 주성분 벡터 $\vec{Y}_t$로 선형변환시킨다. 화자의 주성분 벡터를 이용하여 GMM의 최대 사후확률 값을 갖는 화자모델 n을 찾는다.

$$\hat{S} = \arg\max_{1\leq n\leq S} \sum_{t=1}^{T} \log p(\vec{Y}_t|\theta_n) \quad (22)$$

V. 실험 및 결과

본 논문에서 제안한 방법을 검증하기 위하여 실험에

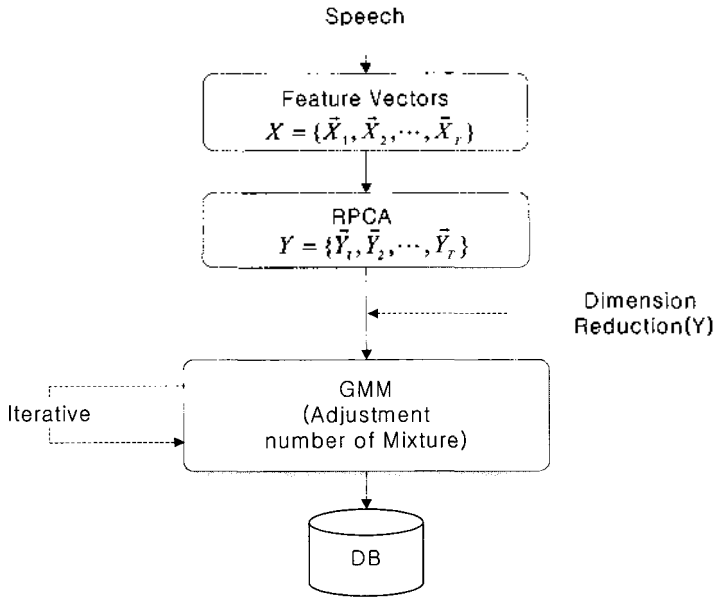


그림 2. 화자 모델을 위한 GMM 학습 과정
Fig. 2. GMM training process for Speaker model.

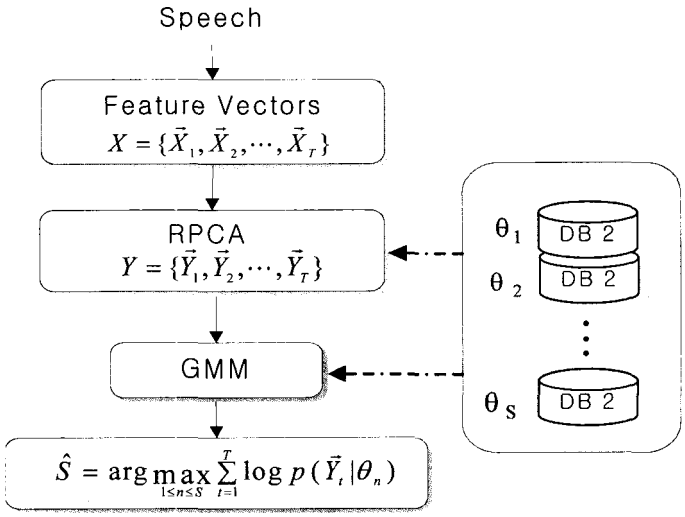


그림 3. 화자식별
Fig. 3. Speaker identification.

사용된 음성 데이터는 200명 (남자 100명, 여자 100명)의 화자가 발성한 한국어 문장 종속 연속음 열려라 참깨 음성이다. 수집된 데이터 음성은 한 화자당 1회에 5번씩 발성한 뒤, 1주 간격의 시간차를 가지고 3주에 걸쳐서 수집하였다. 개인별 전체 발성된 데이터 수는 15개이고 수집된 총 데이터는 3,000개이다. 16 kHz로 샘플링하였고 음성 분석을 위하여 해밍창이 사용하였으며, 한 프레임은 50% 중첩된 256샘플을 사용하였다. 특징 벡터로는 12차 LPC 켑스트럼과 12차 델타 켑스트럼과 델타 에너지를 포함하여 전체 25차를 사용하였다.

학습을 위한 데이터는 각 화자가 2주간 발성한 10개의 음성 데이터를 학습에 사용하였고 마지막 주의 5개의 음성 데이터를 테스트에 사용하였다. 따라서 화자식별 테스트에 사용된 참여 데이터는 1,000개이다.

수집된 음성을 이용하여 제안된 방법의 성능 검증을 위하여 세 가지 실험을 하였다.

첫 번째 실험은 특징벡터의 고차원 문제를 해결하기 위하여 주성분 분석과 강인한 주성분 분석의 차원 수에 따른 화자식별 성능을 살펴보았다. 이 실험에서는 수집된 원음성으로부터 특징벡터를 추출하여 차원수를 11에

서 25차까지 증가시켜 성능을 얻었다 (표 1). 이 실험에 사용된 GMM 혼합 성분 개수는 14개이다. 두 번째 실험은 첫 번째 실험에서 구한 최적의 차원을 사용하여 선형 변환된 특징 벡터를 이용하여 GMM 혼합 성분 개수를 변화시켜 기존의 방법들과 제안된 방법의 성능을 비교한 것이다. 마지막 실험은 음성 검출을 통하여 얻어진 음성 신호와 특징벡터들에 이상치를 추가시킨 다음, 첫 번째와 두 번째에서 선택된 최적의 차원과 GMM 혼합 성분 개수를 이용하여 기존의 알고리즘과 제안된 방법의 outlier문제에 따른 화자식별 성능을 비교하였다. 여기에서 화자인식을 위한 음성 검출은 에너지가 높은 확실한 음성 구간만을 사용하였는데, 이상치의 영향을 실험하기 위하여 검출된 음성구간의 앞뒤 프레임을 증가시키고 특징벡터를 추출하고 또한 임의의 이상치들을 추가시켰다.

표 1은 차원 수에 따른 일반적인 주성분 분석법과 강인한 주성분 분석법의 화자식별 성능을 나타낸 것이다. 화자의 주성분 벡터의 차원이 증가할수록 성능이 향상됨을 알 수 있다. $k \geq 20$일 때는 두 가지 방법 모두 일반적인 GMM ($k = 25$) 방법의 성능보다 높게 나타났다. 즉, 성능의 큰 변화없이 주성분 벡터를 위한 차원을 감소시킬 수 있음을 보여 준다.

그림 4는 화자별 GMM의 혼합성분 개수에 따른 화자식별 성능을 나타낸 것이다. 제안된 방법의 성능은 일반적인 GMM 방법과 비교할 때 평균 0.81%, 주성분 분석법을 사용한 경우 0.06% 더 좋은 화자식별 성능을 보였고, 혼합성분의 개수가 증가할수록 화자식별 성능이 우수함을 보였다. 그러나 혼합성분의 수가 24개 이상인 경우는 혼합성분의 수가 증가해도 일정한 수준의 성능에 접근하면 더 이상 증가하지 않았다. 이것은 학습 데이터가 충분하지 못했기 때문에 나타난 현상이라고 할 수 있다.

표 1과 그림 4는 수집된 원음성에서 추출한 특징벡터를 이용하여 화자식별 성능을 본 것이다. 그 결과 주성분 분석법을 사용한 경우에 일반적인 GMM 방법보다는 성능이 우수하지만 제안된 방법은 일반적인 주성분 분석법과 비

표 1. 차원 수에 따른 화자식별 성능[%]
Table 1. The relationship between the speaker identification rate[%] and k-dimension.

k	PCA	RPCA	k	PCA	RPCA	GMM
10	85	84.6	18	95.2	95.3	
11	89.5	89.6	19	95.9	95.7	
12	90	92.3	20	97	96.6	
13	92.1	92.4	21	97.4	97.4	
14	92.3	92.3	22	97.3	97.4	
15	92.7	92.7	23	97.5	97.7	
16	93.1	93.4	24	97.6	97.9	
17	94.6	94.1	25	98	97.8	95.8

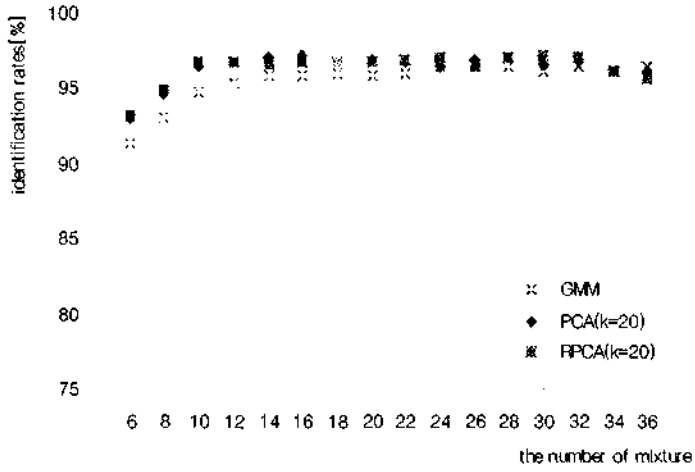


그림 4. 혼합성분 개수에 따른 각 방법의 화자식별 성능[%]
Fig. 4. The relationship between the speaker identification rate[%] and the numbers of mixture.

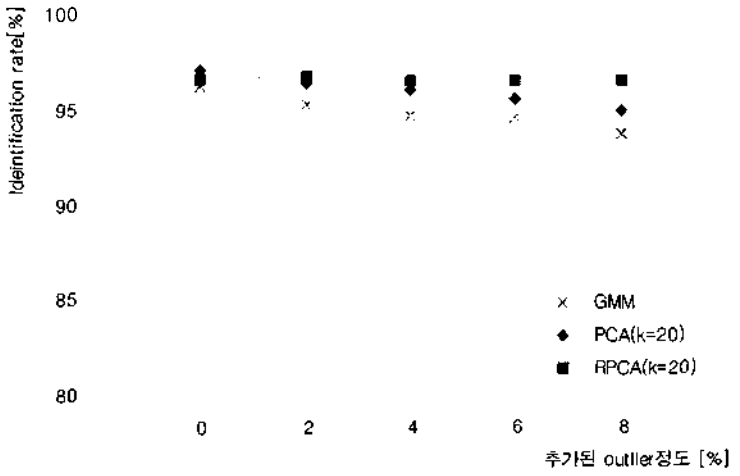


그림 5. Outlier 정도와 화자식별 성능[%]
Fig. 5. The relationship between the speaker identification rate[%] and outliers.

교할 때 비슷한 화자식별 성능을 보였다.

그림 5는 이상치가 존재하는 경우에 GMM, 주성분 분석법, 강인한 주성분 분석법의 성능을 나타낸 것이다. 특징 벡터에 이상치가 존재하는 경우에 화자 모델을 만들어 놓은 다음에 화자식별 성능을 본 것이다. outlier가 2%씩 증가할 때마다 전형적인 GMM 방법과 PCA 방법은 각각 0.65%, 0.55%씩 화자식별 성능이 감소되었지만 제안된 방법은 0.03%정도 감소하였다. 이상치가 증가됨에 따라 제안된 방법의 성능이 일반적인 주성분 분석법을 사용한 경우보다 화자식별률이 높게 나타났지만, 원음성에서 추출된 특징벡터에서는 오히려 제안된 방법의 성능이 일반적인 주성분 분석법보다 성능이 낮게 나타났다. 이상치가 전혀 추가되지 않은 조건에서 제안한 RPCA 방법을 사용할 경우 경계 바깥에 존재하는 특징 벡터들의 영향을 감소시키게 된다. 그러나 특징 벡터들이 경계에 의해 특징 벡터들의 영향이 감소하게 되면 고유의 화자의 특성이 손실을 입게 되므로 기존의 PCA 방법보다 인식 성능이 저하되었다.

이 실험은 수집된 원음성에서 얻은 특징벡터의 경우에는 오히려 강인한 방법을 사용한 경우가 화자식별률에 성능이 감소됨을 알 수 있었다. 그러나 제안된 강인한 주성분 분석법은 이상치 정도가 많이 추가된 경우에는 outlier에 영향을 많이 받지 않고, 성능이 일정하게 유지된 것을 알 수 있다. 즉, 이상치가 존재할 경우에 제안된 방법이 강인함을 보였다. 이는 특징벡터의 파라메타들에 이상치가 존재할 경우 모든 차원 특징벡터를 사용하는 경우와 주성분 분석법에 의해 이상치에 민감하게 반응된 주성분 벡터를 이용할 경우, 화자 식별 성능을 저하 시키게 된다. 그러나 강인한 주성분 분석법을 사용할 경우, 이상치의 영향이 줄어든 특징 벡터로 선형 변환이 되므로 GMM 알고리즘을 이용하여 화자 모델을 구하고, 화자 식별 실험을 한 경우에는 이상치에 강인하게 되는 것이다.

VI. 결 론

본 논문에서는 이상치의 영향과 고차원 문제를 해결하기 위하여 강인한 주성분 분석법을 갖는 GMM 방법을 제안하였다. 음성신호에 이상치가 존재하는 경우 추출된 특징벡터는 이상치에 의한 영향으로 화자의 특성 검출 시 에러를 가져 올 수 있으므로 강인한 주성분 분석법을 이용하였다. 따라서 이상치에 의한 영향을 감소시키고 데이터의 차원을 감소시켜 화자모델을 생성시켜 화자인식 성능을 향상시킬 수 있었다.

또한 제안된 논문은 $k=p$일 때, 기존의 직교 GMM방법으로[2] 대체할 수 있다. 즉 $k=p$일 때, 제안된 방법은 [2]에서 제안된 방법과 동일하므로 [2]의 일반화된 방법이라 할 수 있다.

깨끗한 음성에서 혼합성분 개수에 따라 제안된 방법은 GMM방법과 비교할 때 평균 0.81%, 일반적인 주성분 분석법을 사용한 경우 0.06% 더 좋은 화자식별 성능을 보였다. 이상치가 2%씩 증가할 때마다 일반적인 GMM 방법과 일반적인 주성분 분석법은 화자식별 성능이 급격히 저하되었지만, 제안된 방법의 성능은 깨끗한 음성에서의 성능보다 약 0.1%의 변화만 있었다. 또한 실험 결과에서 제안된 방법은 깨끗한 음성에서보다 약간의 outlier가 존재

할 때 기존 방법과의 큰 성능차이를 보였다.

화자인식의 성능을 더욱 증가시키기 위해서는 시간 경과에 따른 화자 모델의 적응과정이 필요한데 화자모델의 적응과정뿐만 아니라 주성분 분석법의 변환행렬도 적응을 통하여 화자인식의 성능을 더욱 향상시킬 수 있으므로 시간 흐름에 따른 화자별 적응 과정에 대한 연구를 차후 과제로 남겨둔다.

감사의 글

본 논문은 2003학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.

참 고 문 헌

1. D. A. Reynolds, and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Trans. SAP*, 3 (1), 72-83, 1995.
2. L. Liu, and J. He, "On the use of orthogonal GMM in speaker recognition," *ICASSP, Proc.*, 845-849, 1999.
3. C. Seo, K. Y. Lee and J. Lee, "GMM based on local PCA for Speaker Identification," *Electronics Letters*, 37 (24), 1486-1488, 22nd 2001.
4. I. T. Jolliffe, "Principal Component Analysis," New York: Springer-Verlag, 1986.
5. C. Croux, and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence function and efficiencies," *Biometrika*, 87 (3), 603-618, 2000.
6. P. J. Huber, "Robust Statistics," New York: Wiley, 1981.
7. A. Dempster, N. Laird, and D. Doubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, 29, 1-38, 1977.

저자 약력

● 이 윤 정 (Youn-Jeong Lee)

2001년 2월: 숭실대학교 정보통신 공학과 (공학사)
2001년~2003년 2월: 숭실대학교 정보통신 공학과 (석사)
2003년 3월~현재: 숭실대학교 정보통신공학과 박사과정
※ 주관심분야: 화자인식, 음성신호처상, 신경망

● 서 창 우 (Chang-Woo Seo)

한국음향학회지 제22권 제1호 참조
현재: 인스 모바일 기술 연구소 선임 연구원

● 강 상 기 (Sang-Ki Kang)

한국음향학회지 제21권 제4호 참조
현재: 삼성전자 정보통신총괄 통신연구소

● 이 기 용 (Ki-Yong Lee)

한국음향학회지 제15권 제3E호 참조
1997년 9월~현재: 숭실대학교 정보통신전자공학부 부교수