

# PCA 퍼지 혼합 모델을 이용한 화자 식별

## Speaker Identification Using PCA Fuzzy Mixture Model

이 기 용\*

Ki-Yong Lee

### Abstract

In this paper, we proposed the principal component analysis (PCA) fuzzy mixture model for speaker identification. A PCA fuzzy mixture model is derived from the combination of the PCA and the fuzzy version of mixture model with diagonal covariance matrices. In this method, the feature vectors are first transformed by each speaker's PCA transformation matrix to reduce the correlation among the elements. Then, the fuzzy mixture model for speaker is obtained from these transformed feature vectors with reduced dimensions. The orthogonal Gaussian Mixture Model (GMM) can be derived as a special case of PCA fuzzy mixture model. In our experiments, with having the number of mixtures equal, the proposed method requires less training time and less storage as well as shows better speaker identification rate compared to the conventional GMM. Also, the proposed one shows equal or better identification performance than the orthogonal GMM does.

**Keywords :** GMM, PCA, Fuzzy mixture model, Orthogonal, Speaker Identification

### 1. 서 론

가우시안 혼합모델(Gaussian Mixture Model: GMM)(Reynold. 1995, Reynold and Rose. 1995)과 퍼지 c-평균(FCM)(Tran and Wagner. 1999; Tran et al. 1998)에서 나온 퍼지 가우시안 혼합 모델(FGMM)은 화자 식별과 확인의 성능을 높이기 위하여 사용되는 방법들이다. 일반적으로 이러한 모델들은 대각 공분산 행렬로 가정한다. 그리고 높은 차원의 특징 벡터들은 화자 인식기의 성능을 향상 시키기 위해 사용되어 왔다. 그러나 주어진 음성 신호로부터 추출된 특징벡터의 요소들은 많은 상관관계들이 존재하므로, 이러한 상관관계들을 모델링하기 위해서는 많은 혼합 성분들이 필요하다. 또한, 이러한 혼

---

\* 숭실대학교 정보통신전자공학부

합 성분의 개수와 특징벡터의 차원의 증가는 클래스를 나타내기 위해 더 많은 파라미터와 더 많은 저장 공간을 필요로 하고, 많은 양의 데이터는 학습에 소요되는 계산량의 증대를 가져온다.

PCA는 특징벡터의 차원 감소와 특징벡터들 사이에 존재하는 상관 관계를 줄이기 위한 방법 중 하나이다(Jolliffe, 1986). 화자식별에서 PCA를 이용한 직교 GMM 방법은 전형적인 GMM 방법보다 더 좋은 성능을 갖는다(Liu and He, 1999; Seo et al. 2001; 이윤정 외, 2003).

본 논문에서는 화자식별을 위하여 PCA 퍼지 혼합 모델을 제안하였다. 제안된 방법은 대각 공분산 행렬을 갖는 PCA와 혼합 모델의 퍼지설의 결합으로 정의된다. 퍼지 혼합 모델은 FCM의 수정된 FGMM과 구별된다 (Tran and Wagner, 1999; Tran et al. 1998). 제안된 방법에서, 특징벡터는 PCA에 의해 선형변환 된다. 그 결과, 특징벡터들의 요소들 사이에 상관관계가 줄어들게 된다. 화자의 퍼지 혼합 모델은 변환된 특징벡터들로부터 얻어진다. 제안된 방법의 성능을 검증하기 위하여 제안된 방법, 기존의 일반적인 GMM과 직교 GMM 방법의 비교 실험을 통하여 확인하였다.

## 2. PCA 퍼지 혼합 모델

화자의 데이터 집합의 특징 벡터들의 상관관계성과 차원을 감소 시키기 위하여, 각 화자의 PCA를 취하여, 입력 특징 벡터 집합을 시스템의 출력 특징 벡터로 선형 변환시켰다

$x_i$  는  $n$ -차원 특징 벡터이고,  $X = \{x_1, \dots, x_T\}$  는 화자의 특징벡터 집합이라고 가정하자. PCA에서는, 관찰된  $X$  는  $s$ -번째 화자의 선형변환 행렬  $\Phi_s$ 로부터  $L$ -차원 특징벡터  $y_i$ 로 변환 시킨다.

$$y_i = \Phi_s^T (x_i - E[x]) \quad (1)$$

여기에서  $L \leq n$ 이고,  $E[x]$ 는

$$E[x] = \frac{1}{T} \sum_{i=1}^T x_i$$

로 정의된다. 벡터  $\phi_k$  는 공분산 행렬의  $k$ -번째 가장 큰 고유값에 해당되는 고유벡터이고 고유 벡터들로 구성된  $\Phi_s = (\phi_1, \dots, \phi_L)$ 는 변환 행렬이다.

여기에서,  $\Sigma\phi_k = \lambda_k\phi_k$ 이다.  $L$  개의 주성분 축을 갖는  $\Phi_s$ 는 최대 투영이 되는 분산을 갖는 직교축을 갖는다. 식 (1)에서, 각 화자의 변환 행렬  $\Phi_s$ 는 유일하다.

화자 식별을 위하여, PCA 변환된 데이터  $Y = \{y_1, \dots, y_T\}$ 와 혼합 모델의 퍼지설의 조합으로 얻어진 PCA 퍼지 혼합 모델은 다음 식으로 표현된다.

$$b(y_i | \Theta_s) = \left[ \sum_{i=1}^M (w_i p_i(y_i))^\alpha \right]^{1/\alpha} \quad (2)$$

여기에서  $p_i(y_i)$ 는 평균 벡터  $\mu_i$ 와 공분산 행렬  $\Sigma_i^y$ ,  $i$ -번째 선형 조합 계수인  $w_i$ 에 의해서 파라미터화된 가우시안 성분이다. 그리고,  $\alpha$ 는  $\alpha > 0$ 인 퍼지 정도이다.  $p_i(y_i)$ 는 가우시안 함수에 의해서 다음 식으로 나타낼 수 있다.

$$p_i(y_i) = \frac{1}{(2\pi)^{L/2} |\Sigma_i^y|^{1/2}} \exp \left\{ -\frac{1}{2} (y_i - \mu_i)^T \Sigma_i^{y^{-1}} (y_i - \mu_i) \right\} \quad (3)$$

여기에서 선형 조합 계수는  $\sum_{i=1}^M w_i = 1$ 을 만족한다. 변환 행렬  $\Phi_s$ 에 의해 PCA 특징 벡터  $y_i$ 는 직교성을 갖기 때문에 상관관계가 없어 지게 된다.  $\Sigma_i^y$ 는 순수한 대각행렬은 아니지만, 데이터  $X$ 의 공분산 행렬  $\Sigma_i^x$ 보다 대각 성분이 더 우세하다. 그러므로 화자 모델을 위한 PCA 혼합 모델은 평균 벡터, 공분산 행렬, 혼합 성분의 가중치로 파라미터화 된다.

$$\Theta_s = \{\theta_i\}, \quad i = 1, \dots, M, \quad (4)$$

여기에서  $\theta_i = \{w_i, \mu_i, \Sigma_i^y\}$ 이다.

### 3. PCA 혼합 모델을 갖는 화자 모델의 학습 방법

퍼지의 목적은 주어진 변환된 특징 벡터를 위한 PCA 퍼지 목적 함수가 최대인 화자 모델을 찾는것이다. 퍼지 목적 함수는 다음 식과 같다.

$$\begin{aligned} J(Y) &= \sum_{i=1}^T \log \{ b(y_i | \Theta_s) \} \\ &= \sum_{i=1}^T \log \left\{ \sum_{i=1}^M [w_i p_i(y_i)]^\alpha \right\}^{1/\alpha} \end{aligned} \quad (5)$$

제안된 방법의 기본 개념은  $\theta_i$ 가 주어졌을 때  $J$ 를 최대화 하는 것이다.

$$\frac{\partial J}{\partial \theta_i} = \frac{1}{\alpha} \sum_{i=1}^T \left( \frac{(w_i p_i(y_i))^\alpha}{\sum_{m=1}^M (w_m p_m(y_i))^\alpha} \right) \frac{\partial \log(w_i p_i(y_i))}{\partial \theta_i} = 0$$

$J$ 의 최대화는 파라미터  $\theta_i$ ,  $i=1, \dots, M$ 에 대하여  $J$ 의 도함수로부터 얻어지게 되고, 그 결과로 다음 재추정식을 구할 수 있다.

- 혼합 가중치:

$$w_i = \frac{1}{T} \sum_{i=1}^T f_{i,i} \quad (6.a)$$

- 평균 벡터:

$$\mu_i = \frac{\sum_{i=1}^T f_{i,i} y_i}{\sum_{i=1}^T f_{i,i}} \quad (6.b)$$

- 분산행렬:

$$\Sigma_i^y = \frac{\sum_{i=1}^T f_{i,i} (y_i - \mu_i)(y_i - \mu_i)^T}{\sum_{i=1}^T f_{i,i}} \quad (6.c)$$

식(6)에서,  $i$ -번째 성분의 퍼지 소속 함수는 다음 식으로 주어진다.

$$f_{i,i} = \frac{(w_i p_i(y_i))^\alpha}{\sum_{m=1}^M (w_m p_m(y_i))^\alpha} \quad (7)$$

$\alpha=1$ 이고,  $L=n$ 일 때, 직교 GM(Liu and He, 1999)는 PCA퍼지 혼합 모델은 본 논문에서 제안된 모델의 특별한 경우로 간주된다.  $\alpha$ 가 무한대로 가면,  $f_{i,i}$ 는 하드 소속함수가 되고, 전형적인 k-평균 클러스터링 방법과 같게 된다.

#### 4. 화자 식별

화자 식별을 위해서 PCA 퍼지 혼합 모델  $\Theta_1, \Theta_2, \dots, \Theta_S$ 은 화자 특성을 확률 모델의 형태 나타내었다(Reynold and Rose, 1995). 고유한 화자의 속하는 것으로서 의미된 특징벡터에 대해서 PCA 퍼지 혼합 모델은 화자  $S = \{1, 2, \dots, S\}$ 에 대해서 식(2)의 가우시안 요소의 선형 조합으로 이루어진다.

따라서, 고유한 화자 모델은 화자 자신의 특징 벡터에 기초한 PCA 퍼지 혼합 모델에서 적합한 파라미터를 찾는 것이다. 성능 평가를 위한 특징 벡터 열은 식별을 위해서  $N$  개 특징벡터의 중첩된 세그먼트로 나누어진다(Reynold and Rose, 1995; Wang et al. 2002).

벡터열의 첫 번째 두 세그먼트는 다음과 같다.

$$\begin{array}{c}
 \text{Segment } l \\
 \underbrace{y_l, y_{l+1}, \dots, y_{l+N-1}, y_{l+N}, \dots} \\
 \text{Segment } l+1 \\
 \underbrace{y_l, y_{l+1}, \dots, y_{l+N-1}, y_{l+N}, y_{l+N+1}, \dots}
 \end{array}$$

테스트 세그먼트  $Y^{(l)} = \{y_l, y_{l+1}, \dots, y_{l+N-1}\}$  에 대해서, PCA 퍼지 혼합 모델의 로그 유사도 함수는 다음과 같다.

$$\begin{aligned}
 (Y^{(l)}, \Theta_s) &= \sum_{l=1}^{l+N-1} \log b(y_l | \Theta_s) \\
 &= \sum_{l=1}^{l+N-1} \log \left\{ \sum_{m=1}^M [w_m P_m(y_l)]^\alpha \right\}^{1/\alpha} \quad s=1, \dots, S
 \end{aligned} \tag{8}$$

각 세그먼트에서 식별된 화자를  $s^*$  라 하고, 최대 유사도를 갖는 화자를 선택한다.

$$s^* = \arg \max_{1 \leq s \leq S} L(Y^{(l)}, \Theta_s). \tag{9}$$

각 세그먼트에서 식별된 화자는 정확하게 식별된 세그먼트의 테스트 세그먼트 수에 대해서 현재 화자와 비교한다. 최종성능 평가는 정확하게 식별된  $N$ -길이를 모든 테스트 길이로 나눈 백분율로 계산된다.

$$\text{식별율 [\%]} = \frac{\# \text{ correctly identified segments}}{\text{total \# of segments}} \times 100. \tag{10}$$

모든 화자의 테스트 횟수는 거의 같으므로 성능 평가는 거의 공정하게 이루어진다.

## 5. 실험결과

제안된 방법의 효율성을 보여주기 위하여 화자 식별 실험을 하였다. 실험에서 음성 데이터는 4 달 동안 3 회에 걸쳐 50 명(남자 25 명, 여자 25 명)의 화자에 의해서 한국어로 구성되었다. 각 화자는 1 회에 5 번 같은 패스워드를 발성하였다. 11 khz로 샘플링된 음성 데이터는 12 차원 LPC 켈스트럼과 13 차원 델타 켈스트럼을 사용하여 파라미터화하였다. 프레임 길이는 10 ms 중첩된 20 ms를 사용하였다. 1 회에 녹음한 50 개의 문장은 학습에 사용하였다. 2 회와 3 회에 녹음한 250 개의 패스워드를 평가를 위하여 사용하였다.

제안된 방법은 변환 행렬에 대해서 여분의  $L \times n$  저장 공간이 필요하다. 표 1은 기존의 GMM과 제안한 방법에서 화자 식별률과  $\alpha$  값 사이의 관계를 보여준다. 표에서 화자 식별률은  $\alpha$  값에 의해서 변한다.  $\alpha=1.1$  일 때 화자 식별률은 제안된 방법이 좋은 성능을 보였다.

제안된 방법과 기존의 GMM에서 요구되는 파라메타의 수는 표 2에 나타내었다. 화자 모델을 위해서 제안된 방법에서는  $M_p(1+2L)+L \times n$ , 기존 GMM에서는  $M_c(1+2n)$ 의 저장 공간이 필요하지만, 파라메타 수는 제안된 방법이 기존 GMM보다 더 적게 필요하였다.

예를 들어, 같은 성능에서  $M_p=16$ ,  $M_c=32$ ,  $L=15$ ,  $n=25$ 라 할 때, 제안한 방법과 기존 방법은 각각 871 개와 1,632 개의 파라미터를 갖는다. 비록 제안된 방법이 변환 행렬을 얻기 위하여 여분의 단계가 필요하지만, 모든 단계의 계산에 관한 소요시간은 무시할 수 있다. 테스트 부분에서 제안한 방법은 모든 테스트 벡터에 대해서 변환 과정 소요시간이 필요하다. 그러나 작은 계산 오버헤드는 응답 속도에 미치는 영향은 무시할 수 있다.

그림 1은 화자 식별률과 혼합 개수사이의 관계를 보여준다. 그림에서 같은 혼합 개수를 사용하였을 때 제안한 방법이 기존의 GMM과 직교 GMM보다 항상 높은 식별률을 가진다. 같은 성능 레벨에서 기존의 GMM이 제안한 방법에 사용된 혼합 개수의 두 배가 더 필요하였다.

그림 2는 변환된 특징 벡터에 대한 화자 인식률을 보여준다. 그림으로부터  $11 \leq L \leq 25$ 을 가지는 PCA 방법이  $n=25$ 를 가지는 기존의 GMM보다 더 좋음을 알 수 있다. 그리고 제안한 방법이 직교 GMM 방법보다 같거나 더 좋은 식별 성능을 보였다.

## 6. 결 론

화자 식별을 위하여 PCA 퍼지 혼합 모델을 제안하였다. 각 화자의 변환 행렬을 이용하여 차원을 줄였고, PCA 퍼지 혼합 모델은 PCA와 대각 공분산 행렬을 가지는 혼합의 퍼지 버전과 결합으로 정의된다.  $\alpha = 1$  일 때, 제안한 방법은 직교 GMM의 특수한 경우가 된다.

실험결과에서 같은 클러스터 개수를 가졌을 때 제안한 방법이 기존의 GMM과 비교하였을 때 학습시간과 저장 공간이 덜 필요할 뿐만 아니라 화자 식별률이 더 좋음을 알 수 있었다. 또한, 제안한 방법은 직교 GMM 방법보다 같거나 더 좋은 식별 성능을 보였다.

## 참 고 문 헌

- [1] Seo, C., Lee, K.Y. and Lee, J. 2001. "GMM based on local PCA for speaker identification." *IEEE Electronic Letters*, 37 (24), 1486-1488.
- [2] Reynolds, D.A. 1995. "Speaker identification and verification using Gaussian mixture speaker models." *Speech Comm.*, 17, 91-108.
- [3] Reynolds, D.A. and Rose, R.C. 1995. "Robust text-independent speaker identification using Gaussian mixture speaker models." *IEEE Tr. SAP*, 3(1), 72-82.
- [4] Tran, D. and Wagner, M. 1999. "A robust clustering approach to fuzzy Gaussian mixture models for speaker identification." *Proc. KES'99*, Adelaide, Australia, 337-340.
- [5] Tran, D., Le, T.V, and Wagner, M. 1998. "Fuzzy Gaussian mixture models for speaker recognition." *Proc. ICSLP*, 3, 759-762.
- [6] Jolliffe, I.T. 1986. "Principal component analysis," *Springer-Verlag*.
- [7] Liu, L. and He, J. 1999. "On the use of orthogonal GMM in speaker recognition." *Proc. ICASSP*, 845-849.
- [8] Wang, L., Chen, K. and Chi, H.S. 2002. "Capture interspeaker information with a neural network for speaker identification." *IEEE Tr. Neural Network*, 13(2), 436-445.
- [9] 이윤정, 서창우, 강상기, 이기용. 2003. "화자식별을 위한 강인한 주성분 분석 가우시안 혼합 모델." *한국음향학회지*, 22(7), 519~527.

접수일자: 2003. 11. 15.

게재결정: 2003. 12. 20.

## ▲ 이기용

서울시 동작구 상도 5동 1-1 (우: 156-743)

송실대학교 정보통신 전자공학부

Tel : +82-2-820-0908 Fax: +82-2-817-4591

E-mail: kylee@ssu.ac.kr



표 1.  $\alpha$  와 화자식별률과의 관계 ( $M=32, L=25$ )

$\alpha$	화자식별률 [%]	
	PCA를 하지 않는 제안된 방법	PCA를 하는 제안된 방법
0.7	91.88	93.13
0.8	91.50	92.50
0.9	92.13	92.88
1.0	92.38	92.90
1.1	92.50	93.25
1.2	92.75	93.21

표 2. 제안된 방법과 전형적인 GMM에 필요한 파라메타

Proposed Method	Conventional GMM
$M_p(1+2L)+L \times n$	$M_c(1+2n)$

$M_p$  : 제안된 방법에서의 혼합성분 개수

$M_c$  : GMM에서의 혼합성분 개수

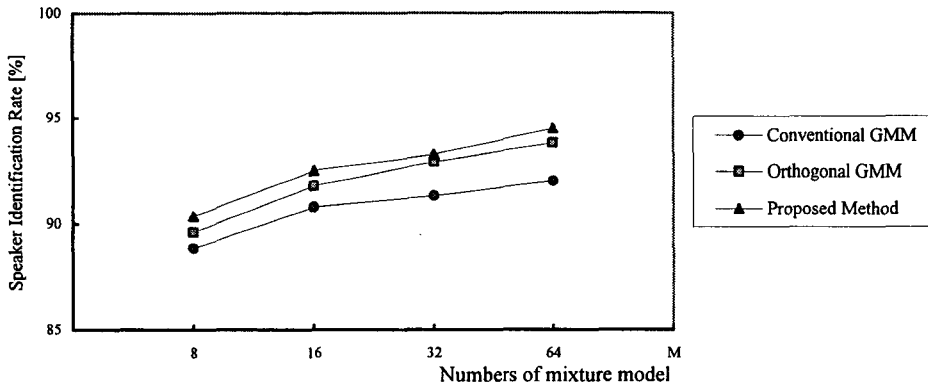


그림 1. 혼합성분 개수와 화자 식별률 사이의 관계 ( $\alpha = 1.1, L=25$ )

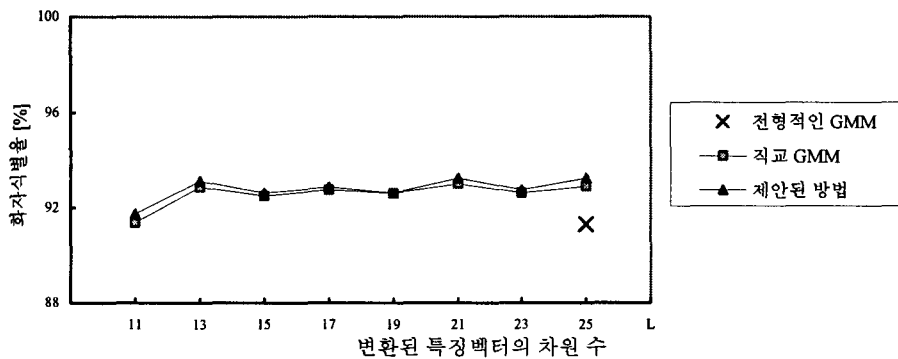


그림 2. 변환된 특징벡터의 차원수와 화자 식별률과의 관계 ( $\alpha = 1.1, M=32$ )