

LPCA에 기반한 GMM을 이용한 화자 식별

Speaker Identification Using GMM Based on LPCA

서창우* · 이윤정** · 이기용**

Changwoo Seo · Younjeong Lee · Ki Yong Lee

ABSTRACT

An efficient GMM (Gaussian mixture modeling) method based on LPCA (local principal component analysis) with VQ (vector quantization) for speaker identification is proposed. To reduce the dimension and correlation of the feature vector, this paper proposes a speaker identification method based on principal component analysis. The proposed method firstly partitions the data space into several disjoint regions by VQ, and then performs PCA in each region. Finally, the GMM for the speaker is obtained from the transformed feature vectors in each region. Compared to the conventional GMM method with diagonal covariance matrix, the proposed method requires less storage and complexity while maintaining the same performance requires less storage and shows faster results.

Keywords: Speaker Identification, PCA, VQ, GMM

1. 서론

개인의 고유한 정보를 포함하고 있는 음성 신호를 이용하여 각 개인의 화자 정보를 추출하여 개인을 확인하는 것을 화자 인식이라고 한다. 화자 인식은 화자 식별과 화자 확인으로 나눌 수 있다. 화자 식별은 발생된 음성 신호가 등록된 화자들 중에서 어떤 화자인지를 골라내는 것이고, 화자 확인은 발생된 음성 신호가 등록된 화자의 음성과 일치하는지를 판정하는 것으로, 발생한 화자와 등록된 화자와의 확인과정을 통하여 문턱값보다 유사도가 큰 경우 수락하고, 작은 경우 거절하는 것이다. 본 논문에서는 화자 인식 중에서 화자 식별의 성능 향상을 위한 방법을 제안하고자 한다. 음성신호로부터 개인 정보를 추출하기 위하여, 현재 가장 널리 사용되

* ㈜인스모바일 기술연구소

** 숭실대학교 정보통신공학과

고 있는 분석은 LPCC(linear prediction cepstral coefficients)와 MFCC(mel frequency cepstral coefficients) 방법 등이 있다. 그러나 이러한 일반적인 특징벡터는 화자인식 시스템에서 신호의 상관성 때문에 성능감소가 일어난다[1][2]. 따라서 높은 인식률을 얻기 위해서는 많은 파라미터 개수를 필요로 하지만, 파라미터의 증가는 많은 계산량을 필요로 하기 때문에 실시간 구현을 보다 어렵게 만든다. 그러므로, 특징벡터의 차원을 감소시키기 위해서 주성분 분석을 이용한 화자인식 방법이 제안되었다[3][4][5][6]. 주성분 분석(principal component analysis)은 원 신호의 특징벡터를 선형변환을 통하여 특징벡터의 상관성을 줄이고, 차원을 감소시키는 방법이다.

본 논문에서는 화자 식별을 위하여 특징 벡터의 차원을 효과적으로 감소시키기 위해, 벡터 양자화(Vector Quantization: VQ)에 의한 국부 주성분 분석을 이용한 GMM 방법을 제안한다. 제안된 방법은 먼저 VQ를 통하여 특징벡터 공간을 여러 개의 분리(disjoint)된 공간으로 나눈다[7]. 둘째, 각 부영역에서 주성분 분석을 적용하여 감소된 차원을 갖는 특징 벡터를 구한다. VQ에 의한 각 특징벡터의 클러스터링에서 공분산 행렬을 구하고, 이로부터 고유 벡터를 추출하여 $L(<n)$ 차원을 갖는 선형 부분 공간으로 특징 벡터를 선형 변환 시킨다. 마지막으로 대각 공분산 행렬을 갖는 GMM이 변환된 특징 벡터들로부터 얻어진다. 제안된 방법의 성능을 확인하기 위하여 전형적인 GMM 방법과 제안된 방법을 화자인식을 통하여 성능을 비교한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 VQ에 의한 국부 주성분 분석에 대하여 제안하고, 3장에서는 제안된 방법의 성능을 확인하기 위한 실험결과를 설명한다. 마지막 4장에서 결론을 내린다.

2. 국부 주성분 분석을 이용한 GMM

화자식별을 위하여 특징벡터의 차원을 줄이기 위해서 주성분 분석이 제안되었다. 주성분 분석은 원 신호의 특징벡터를 선형변환을 통하여 작은 부분공간으로 사영시킴으로써 특징벡터의 차수와 상관성을 줄이는 방법이다. 본 논문에서는 벡터 양자화(VQ)를 통한 국부 주성분 분석을 유도하여 GMM에 효과적으로 적용할 수 있는 알고리즘을 제안한다.

제안된 방법에서 첫 번째 과정은 벡터 양자화를 통한 특징 공간을 몇몇 영역으로 분리시키는 것이다. 두 번째 과정은 각 영역에서 PCA를 통한 차원이 줄어든 새로운 특징벡터를 얻는 것이다. 이때 주성분 분석은 각 영역에서 특징벡터의 공분산 행렬의 고유벡터를 통한 n 차원 특징벡터를 L 차원 선형 부분공간으로 변환한다. 마지막으로, 대각 공분산 행렬(diagonal covariance matrix)을 갖는 GMM이 변환된 특징벡터로부터 얻어진다. 각 부분의 의미와 수식에 관한 내용이 다음 절에서 설명된다.

2.1 벡터 양자화를 통한 국부 주성분 분석

벡터 양자화를 이용한 데이터 압축 기법은 단순하고 압축률이 높기 때문에 음성이나 영상의 데이터 압축 방법으로서 지난 수십 년간 많은 관심을 받아왔다. 벡터 양자화는 두 단계로 구성된다. 먼저 부호화 단계에서는 주어진 표준벡터 집합에서 입력 벡터에 가장 가까운 대표 벡터의 하나를 찾아내어 대표 벡터의 색인을 전송하는 과정이고, 복호화 단계는 전송된 색인에 해당하는 대표 벡터를 찾아내는 과정이다.

여기서 중요한 문제는 주어진 학습 데이터를 사용해 어떻게 이 학습 데이터를 가장 잘 대표할 수 있는 대표 벡터 집합인 코드북을 구하느냐에 관한 문제이다. 즉, 주어진 학습 데이터 각각에 대해서 이것을 어느 한 코드북으로 근사화 했을 때 그것들 사이의 평균 거리오차가 최소가 되는 코드북을 구하는 것이다. 이러한 코드북을 구하는 알고리즘으로 여러 가지가 제안되었지만 그 중에서 K 평균 알고리즘이 가장 널리 사용되고 있다.

K 평균 알고리즘의 목적은 주어진 학습 벡터 $X = \{x_1, \dots, x_T\}$ 에 대해서 식 (1)의 목적 함수 J 를 최소로 하는 분할 $W = \{w_{ij}, 1 \leq i \leq T, 1 \leq j \leq K\}$ 와 코드북 $Z = \{z_j, 1 \leq j \leq K\}$ 를 구하는 과정으로 파악할 수 있다.

$$J = \sum_{i=1}^T \sum_{j=1}^K w_{ij} d(x_i, z_j) \quad (1)$$

여기서 w_{ij} 는 0 또는 1의 값을 가지며 제곱 평균 거리오차 $d(x_i, z_j)$ 는 다음과 같이 구할 수 있다.

$$d(x_i, z_j) = \frac{1}{n} \sum_{k=1}^n (x_{ik} - z_{jk})^2 \quad (2)$$

여기서 n 은 벡터 차원이다. K 평균 알고리즘의 구성은 다음과 같다.

단계 1. 초기 코드북 생성: 초기 코드 벡터의 집합을 설정한다.

$$Z = \{z_j(t), 1 \leq j \leq K\}, \text{ where } t=0 \quad (3)$$

단계 2. 코드워드 분류: 입력 데이터를 주어진 코드북과의 거리를 계산하여 최소거리를 갖는 코드워드로 분류한다.

$$x \in C_i(t), \text{ if } d(x, z_i(t)) < d(x, z_j(t)), \text{ for all } i \neq j \quad (4)$$

단계 3. 코드 벡터 갱신: 각 코드 벡터로 분류된 학습 데이터에 대해서 평균을 구하고 이를 새로운 코드북으로 갱신한다.

$$z_i(t+1) = \text{cent}(C_i(t)), \quad 1 \leq i \leq K, \quad (5)$$

$$t = t + 1$$

단계 4. 종료: 만일 전체 왜곡차 $D(t) - D(t-1) < \varepsilon$ 가 문턱값 ε 보다 작으면 종료하고, 그렇지 않으면 단계 2부터 반복한다.

$$D = \frac{1}{T} \sum_{t=1}^T d[x_i, z_j(t) | x_i \in C_j(t-1)], \quad 1 \leq j \leq K \quad (6)$$

먼저 x_t 는 n 차원 특징벡터이고 $X = \{x_1, \dots, x_T\}$ 는 관측된 데이터 집합이라고 가정한다. 특징벡터 집합 X 는 유클리안 거리를 기본으로 하는 벡터 양자화에 의해서 몇몇 공간으로 분리될 수 있다. 이때, 양자화기는 K 기준벡터의 집합 $z^j, j=1, \dots, K$ 과 대응 영역 R^j 로 구성된다. 기준벡터는 일반적인 k -평균 알고리즘으로 쉽게 얻을 수 있다. 각 영역 R^j 는 어떤 다른 기준벡터보다 z^j 에 근접하게 된다. 식(7)은 각 영역에서의 특징벡터들의 공분산 행렬을 계산하는 것이다.

$$\Sigma_j = E[(x - z^j)(x - z^j)^T | x \in R^j], \quad j=1, \dots, K \quad (7)$$

위 식으로부터 추정된 공분산 행렬은 고유치와 대응하는 고유벡터 행렬을 구하기 위해서 사용된다. 어떤 변환된 축의 중요성은 고유치의 크기에 의해서 결정되기 때문에 가장 큰 고유치와 그 대응하는 L 차원의 중요한 고유벡터가 선택되었다. 이때, 고유벡터는 원 신호를 최적으로 나타내기 위해서 특징벡터의 변환에 사용되었다.

그러므로, 학습과 테스트에서 각 입력 특징벡터는 다음과 같이 변환되었다.

$$y_{i,j} = \Omega_j^T x_i, \quad \text{if } x_i \in R^j, \quad 1 \leq j \leq K \quad (8)$$

여기서 Ω_j^T 는 R^j 의 L 차원의 중요한 고유벡터인 행(row)을 갖는 $n \times L$ 가중치 행렬이다. 식 (8)의 공분산 행렬은 대각 행렬 형식을 가진다.

2.2 VQ-GMM 모델 설명

화자인식 시스템에서 특징 파라미터를 사용하는 GMM의 출력확률은 특징벡터 각각의 관측 벡터에 대한 확률 곱의 형태로 나타낼 수 있다. 국부 주성분 분석을 통한 특징벡터의 j 번째 영역에서 T_j 학습벡터 열 $Y_j = \{y_{i,j=1}, \dots, y_{i,j=T_j}\}$ 이 주어질 때, 가우시안 mixture 밀도는 다음과 같이 M_j 성분 밀도의 가중화된 합으로 정의된다.

$$p(y_{i_j} | \lambda) = \sum_{m=1}^{M_j} p_{j_m} b_m(y_{i_j}) \quad (9)$$

여기서

$$b_m(y_{i_j}) = \frac{1}{(2\pi)^{\frac{L}{2}} |\Sigma_{j_m}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y_{i_j} - \mu_{j_m})^T \Sigma_{j_m}^{-1} (y_{i_j} - \mu_{j_m})\right\} \quad (10)$$

는 평균 μ_{j_m} 와 분산 Σ_{j_m} 을 가지는 가우시안 확률밀도 함수를 나타내고, M_j 는 j 영역의 mixture 개수이며, mixture 가중치 p_{j_m} 은 다음 조건을 만족시켜야 한다.

$$\sum_{j=1}^K \sum_{m=1}^{M_j} p_{j_m} = 1 \quad (11)$$

선형 변환된 특징벡터 $Y = \{Y_1, \dots, Y_K\}$ 가 주어질 때, 화자 모델을 위한 GMM은 모든 성분 밀도로부터 평균벡터, 공분산 행렬 그리고 mixture 가중치로 파라미터화 된다. 이런 파라미터 λ 는 다음과 같이 표기될 수 있다.

$$\lambda = \{p_{j_m}, \mu_{j_m}, \Sigma_{j_m}\}, \quad \begin{array}{l} j = 1, \dots, K \\ m = 1, \dots, M_j \end{array} \quad (12)$$

2.3 최대 유사도를 위한 파라미터 추정

화자로부터 학습 데이터가 주어질 때, 학습 모델의 궁극적인 목적은 학습 데이터의 특징벡터 분포를 가장 잘 매칭시킬 수 있는 GMM의 파라미터 λ 를 추정하는 것이다. GMM의 파라미터를 추정하기 위한 몇 가지 기법이 연구되었다. 지금까지 가장 널리 사용되고 그리고 좋은 추정 결과를 얻을 수 있는 방법은 최대 유사도(ML) 추정이다.

ML 추정의 목적은 학습데이터가 주어질 때, GMM의 유사도를 최대화시키는 모델 파라미터를 구하는 것이다. 이때, GMM 유사도는 다음과 같이 나타낼 수 있다.

$$p(y | \lambda) = \prod_{i=1}^T p(y_{i_1} | \lambda) \cdots \prod_{i_k=1}^{T_k} p(y_{i_k} | \lambda) \quad (13)$$

그러나, 이런 표현은 파라미터 λ 의 비선형 함수에 의한 직접적으로 최대화를 시킬 수 없다 따라서 ML 파라미터 추정은 EM(expectation-maximization) 알고리즘을 사용하여 반복적으로 얻을 수 있다. EM 알고리즘의 기본적인 방법은 초기 모델 λ 로 시작해서 새로운 모델 $\hat{\lambda}$, $p(y | \hat{\lambda}) \geq p(y | \lambda)$ 를 추정하는 것이다. 새로운 모델은 다음 반복에서는 초기 모델로 사용되고 그리고 이런 반복 과정이 수렴할 때까지 계속된다.

EM 알고리즘의 진행과정은 다음과 같다. 먼저 mixture의 개수를 미리 정해놓고 모든 데이터를 초기 개수로 랜덤하게 또는 임의의 클러스터링 기법을 이용하여 구한다. 초기 모드에서는 이렇게 나누어진 클러스터내의 데이터를 이용하여 파라미터를 구한다. 각 클러스터내의 특징벡

터의 부분은 가중치를 두게 되는데 여기서 클러스터의 특징벡터는 가우시안 분포인 경우 평균, 분산으로 나타난다. 이러한 과정이 흔히 EM 알고리즘의 E 단계가 된다.

E 단계 후에는 추정된 Mixture 모델로부터 최대 유사도를 통한 특징벡터의 재 클러스터링이 이루어지게 된다. 이 과정을 M 단계라 한다. 이러한 EM 학습 알고리즘은 국부 최대치 수렴을 보장한다.

각 EM 반복에서 다음 재추정 식은 모델의 유사도를 단조증가 시키기 위해서 사용된다.

- Mixture weights :

$$\hat{p}_{jm} = \frac{1}{T} \sum_{i=1}^{T_j} p(j, m | y_i, \lambda) \quad (14)$$

- Means :

$$\hat{\mu}_{jm} = \frac{\sum_{i=1}^{T_j} p(j, m | y_i, \lambda) y_i}{\sum_{i=1}^{T_j} p(j, m | y_i, \lambda)} \quad (15)$$

- Variances :

$$\hat{\Sigma}_{jm}^2 = \frac{\sum_{i=1}^{T_j} p(j, m | y_i, \lambda) (y_i - \mu_{jm})(y_i - \mu_{jm})^T}{\sum_{i=1}^{T_j} p(j, m | y_i, \lambda)} \quad (16)$$

j 번째 영역에서 mixture m 의 사후확률(a posteriori probability)은 다음과 같이 구할 수 있다.

$$p(j, m | y_i, \lambda) = \frac{p_{jm} b_j(y_i)}{\sum_{k=1}^M p_{jk} b_k(y_i)} \quad (17)$$

여기서 $n=L$ 그리고 $K=1$ 일 때, 제안된 방법은 직교 GMM의 특별한 경우가 된다.

2.4 화자 식별

화자식별(speaker identification)은 발생된 음성 신호가 등록되어 있는 화자들 중에서 가장 유사도가 높은 화자를 골라내는 것이다. S명의 화자로부터, 화자 각각은 LPCA에 기반한 GMM의 $\lambda_1, \lambda_2, \dots, \lambda_s$ 로 나타내고, 화자의 각 국부 주성분 벡터를 이용하여 GMM의 최대 사후확률 값을 갖는 화자모델 n 을 찾는다.

$$\hat{S} = \arg \max_{1 \leq n \leq S} \prod_{i=1}^{T_1} p(y_i | \lambda_n) \cdots \prod_{k=1}^{T_K} p(y_k | \lambda_n) \quad (18)$$

전체 테스트한 음성 신호와 정확한 화자를 찾은 개수를 이용하여 화자 식별률(%)을 계산할 수 있다.

$$\text{Identification}[\%] = \frac{\# \text{correctly identified speaker}}{\text{total \# of speakers}} \quad (19)$$

3. 실험 결과

제안된 LPCA에 기반한 GMM 방법의 성능 확인을 위하여 연구실 환경에서 수집한 한국어 고립단어 숫자음, “영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구”를 사용하였다. 실험에 사용된 화자는 남자 100 명, 여자 100 명으로 총 3 주에 1 주에 한번 방문하여 5 번씩 발성한 숫자음이다. 첫 번째, 실험에서 사용한 학습 데이터는 처음 1 주간 5 회 발성한 데이터를 사용하였고 테스트 데이터는 마지막 2 주에 걸쳐서 수집한 10 회 발성한 데이터를 사용하였다. 마지막 실험에서는 학습 데이터에 처음 2 주간 발성한 10 회 발성 데이터, 그리고 테스트 데이터는 마지막 5 회 발성 데이터를 사용하였다. 샘플링 주파수는 16 kHz를 사용하였고, 한 프레임의 크기는 50 % 중첩을 가진 256 샘플을 사용하였으며, 음성 특징 벡터로는 멜 캡스트럼 25 차를 사용하였다.

알고리즘 검증을 위하여 숫자음을 이용한 문장종속(text-dependent) 화자식별을 적용하였다. 표 1은 주성분 분석의 차수별 고유치와 정보율을 나타낸 것이다. 주성분 분석을 이용하여 선형 변환 시킬 때 원본 데이터의 몇 %의 정보율을 가지고 변환할 것인지를 결정하는데 필요하다. L 차원 주성분 벡터의 정보율(I)은 고유값 Λ 를 이용하여, 다음 식에 의해 구할 수 있다.

$$I[\%] = \frac{\sum_{i=1}^L \Lambda_i}{\sum_{i=1}^n \Lambda_i}$$

그림 1은 $K=1$ 이고, mixture의 개수가 16 개일 때, 주성분 분석의 차수별 정보율이 80% 이상 되는 경우에 화자 식별 성능을 나타낸 것이다. 표 1과 그림 1로부터 특징벡터의 차원 감소를 위해서 25 차 캡스트럼의 정보율이 95% 이상이 되는 차원을 선택하여 95.56%와 100%에 해당하는 15 차와 25 차 캡스트럼에서 실험하였다. 또한 국부 주성분 분석을 이용한 GMM이 효과적인 알고리즘임을 검증하기 위해서 선형 부분공간 VQ의 개수는 $K=1$ 과 $K=2$ 에서 실험을 하였다.

제안된 방법과 일반적인 GMM에서의 필요 파라미터 수는 표 2에 나타나있다. 여기서 M_p 와 M_c 는 제안된 방법과 일반적인 GMM에서의 mixture 개수이다. n 과 L 은 각각 일반적인 GMM에 사용한 25 차 특징벡터와 제안한 방법에서의 15 차 특징벡터를 나타낸다. 여기서 $K=1$ 이고 $L=n$ 일 때, 제안된 방법은 [2]의 특별한 경우로 직교 GMM이라고 할 수 있다.

표 1. 주성분 분석의 차수별 고유치와 정보율(%)

차 수	고 유 치	정보율 (%)	차 수	고 유 치	정보율 (%)
1	55.185	23.184	14	2.534	94.771
2	35.739	39.199	15	1.884	95.563
3	24.981	48.693	16	1.826	96.330
4	19.006	56.678	17	1.665	97.030
5	16.461	63.593	18	1.570	97.689
6	14.684	69.762	19	1.492	98.316
7	12.252	74.909	20	1.340	98.879
8	12.057	79.975	21	1.223	99.393
9	10.812	84.517	22	1.024	99.823
10	8.327	88.016	23	0.362	99.975
11	6.410	90.709	24	0.055	99.998
12	4.318	92.522	25	0.006	100
13	2.819	93.707			

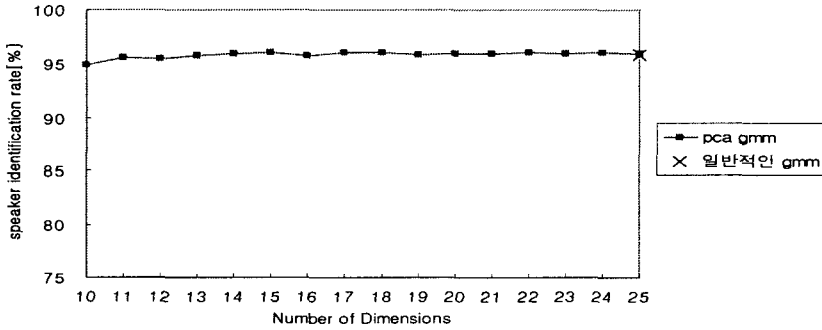


그림 1. 주성분 분석의 차수별 화자인식률 (K=1, mixture:16)(%)

표 2. 제안된 방법과 일반적인 GMM을 위한 필요 파라미터 개수

제안된 GMM	일반적인 GMM
$K \cdot M_p (2L+1) + K \cdot n(L+1)$	$M_c (2n+1)$

M_p : 제안된 GMM에서의 mixture 개수

M_c : 일반적인 GMM에서의 mixture 개수

그림 2는 mixture의 개수를 증가시키면서 화자인식 성능을 본 것이다. 또한 테스트에 사용된 데이터는 학습에 사용을 하지 않은 Open Set에서 실험되었다. 실험 결과로부터 제안한 국부

주성분 분석을 이용했을 때, 선형 부분공간 $K=2$ 이고 특징벡터 차원이 25일 때, 가장 좋은 결과를 보였다. 이 경우 기존의 일반적인 GMM보다 1.5% 이상 좋은 결과를 보였다. 또한 국부 주성분 분석에서 $K=2$ 이고 특징벡터가 $L=15$ 일 때, 일반적인 GMM보다 1% 이상 좋은 실험 결과를 보였다. 선형 부분 공간 $K=1$ 이고 $L=n$ 일 때, 일반적인 GMM보다 0.5% 이상 좋은 결과를 보였다.

제안한 방법이 각각 변환 행렬을 위해서 $K \times L \times n$ 와 $K \times n$ 의 계산량 그리고 VQ를 필요로 하지만, 일반적인 GMM보다 훨씬 작은 파라미터를 요구한다. 제안한 방법에서 $K=2$ 이고 $L=15$ 일 때, 제안한 방법에서의 mixture $M_p=24$ 이고 일반적인 GMM $M_c=48$ 일 때, 제안한 방법과 일반적인 GMM의 필요 파라미터 수는 각각 2288개와 2448개이다. 이 경우 제안한 방법이 일반적인 GMM에서 24 개의 mixture를 사용할 때보다 더 좋은 결과를 보였다.

그림 2가 테스트에 학습 데이터를 참가시키지 않은 반면, 그림 3은 학습 데이터를 참가시킨 Close Set이라고 할 수 있다. 따라서, 이 경우 실험 결과는 학습 데이터를 포함했기 때문에, 전체 테스트 데이터는 3,000개이다. 그림 3에서는 일반적인 GMM이 평균 95.4%인 반면, 제안한 방법에서 $K=1$ 이고 $L=25$ 일 때 96.06%로 신호의 상관성을 제거하는데 주성분 분석이 효과적인 알고리즘임을 확인할 수 있다. 또한 $K=2$ 이고 $L=15$ 와 25일 때, 각각 96.39%와 97.06%가 나왔다.

그림 2와 3을 통해서 알 수 있는 것은 mixture의 개수가 증가하지만 어느 정도 성능이 향상되면 더 이상 증가하지 않은 것을 알 수 있다. 이것은 학습 데이터가 충분하지 못했기 때문에 나타난 현상이라고 할 수 있다. 이런 문제를 줄이기 위해서 그림 4에서는 지금까지의 학습 데이터를 두 배로 증가시킨 경우에 대하여 실험하였다.

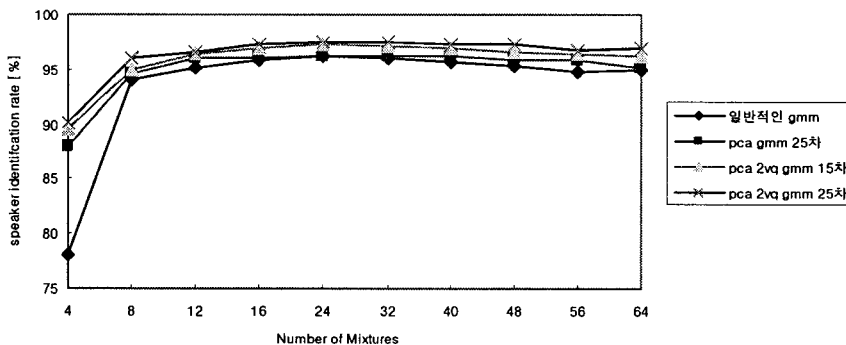


그림 2. Open Set에서의 일반적인 GMM과 제안한 GMM의 성능 비교

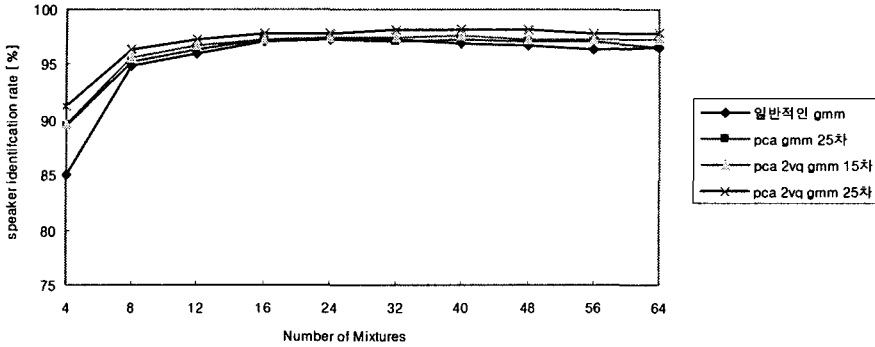


그림 3. Closed Set에서의 일반적인 GMM과 제안한 방법의 성능 비교

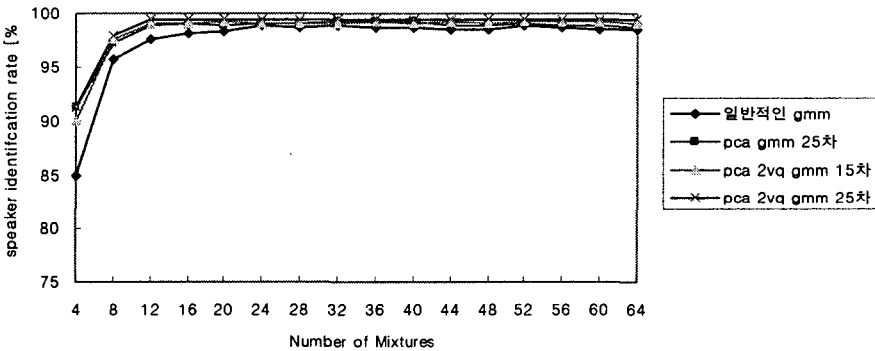


그림 4. 학습 데이터 증가에서의 일반적인 GMM과 제안한 방법에서의 성능 비교

그림 4는 그림 2와 3에서 나타난 문제를 해결하고 제안한 방법이 좋은 실험 결과를 얻을 수 있음을 증명하기 위한 실험이다. 실험은 테스트 데이터를 참여시키지 않은 Open Set에서 실험을 수행하였다. 실험 결과에서 제안한 국부 주성분 분석을 이용했을 때, 제안한 방법에서는 기존의 일반적인 GMM보다 1.3% 이상 좋은 98.33%의 결과를 보였다. 또한 국부 주성분 분석에서 $K=2$ 이고 특징벡터가 $L=15$ 일 때, 일반적인 GMM보다 0.8% 이상 좋은 실험 결과를 보였다. 일반적인 GMM과 제안한 방법에서 $K=1$ 이고 $L=n$ 일 때, 97.52와 98.39%가 보였다.

제안한 방법에서 $K=2$ 이고 $L=25$ 일 때, 그림 2와 3을 비교할 때, 5개의 학습 데이터를 사용한 경우 그림 1은 mixture 24개 이상 증가했을 때, 성능 향상보다는 성능 감퇴가 일어났다. 반면에 학습 데이터가 10번 발생으로 이루어진 그림 3에서는 약간의 성능 향상이 일어나고 있

지만, 성능 감쇠는 거의 일어나지 않는다.

4. 결 론

본 논문에서는 화자 인식에서 벡터 양자화를 통한 국부 주성분 분석을 이용하여 효과적으로 GMM을 모델링 하는 방법을 제안 하였다. 제안된 방법은 데이터 공간을 VQ를 통하여 몇 개의 분리된 클러스터링으로 나누고, 각각의 클러스터에서 PCA를 수행한다. 전형적인 GMM과 비교했을 때, 제안된 방법은 똑같은 성능을 유지하면서, 더 적은 기억 공간을 필요로 하고, 더 빠른 결과를 보였다. 또한, 제안한 방법이 비록 계산과정에서 선형변환과 VQ를 수행해야 된다는 복잡성을 가지고 있지만, 작은 파라미터로 우수한 성능 결과를 얻을 수 있다는 것을 알 수 있다.

감사의 글

본 논문은 2004학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.

참 고 문 헌

- [1] Reynolds, D. and Rose, R. 1995. "Robust text-independent speaker identification using Gaussian mixture speaker models." *IEEE Trans. on SAP*, vol. 3, no. 1, pp. 72-82.
- [2] Liu, L. and He, J. 1999. "On the use of orthogonal GMM in speaker recognition." *International Conference on Acoustics, Speech, and Signal Processing*, pp. 845-849.
- [3] Ariki, Y., Tagashira, S. and Nishijima, M. 1996. "Speaker recognition and speaker normalization by projection to speaker subspace." *International Conference on Acoustics, Speech, and Signal Processing*, pp. 319-322.
- [4] Kambhatla, N. and Leen, T. K. 1997. "Dimension reduction by local PCA." *Neural Computation*, 9, pp. 1493-1503.
- [5] Seo, C. W., Lee, K. Y. and Lee, J. 2001. "GMM based on Local PCA for Speaker Identification." *Electronics Letters* 37, 24, pp. 1486-1488.
- [6] 이윤정, 서창우, 강상기, 이기용. 2003. "화자식별을 위한 강인한 주성분 분석 가우시안 혼합 모델." *한국 음향학회*, vol. 22, no.7, pp. 519-527.
- [7] Gersho, A. and Gray, R. M. 1992. "Vector quantization and signal compression." *Kluwer Academic Publishers*, Norwell, MA.

접수일자: 2004. 11. 22

게재결정: 2005. 05. 30

▲ 서창우

서울시 동작구 상도 5동 1-1

(주)인스 모바일 기술 연구소 (우: 156-743)

Tel : +82-2-817-4591, Fax: +82-2-817-4591

E-mail: cwseo@insmobile.com

▲ 이윤정

서울시 동작구 상도 5동 1-1

승실대학교 정보통신 전자공학부 (우: 156-743)

Tel : +82-2-817-4591, Fax: +82-2-817-4591

E-mail: yjlee@ctsp.ssu.ac.kr

▲ 이기용

서울시 동작구 상도 5동 1-1

승실대학교 정보통신 전자공학부 (우: 156-743)

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

E-mail: kylee@ssu.ac.kr