말소리와 음성과학 제 1 권 제 4 호 (2009)
pp. 127~132

화자확인에서 특징벡터의 순시 정보와 선형 변환의 효과적인 적용

Effective Combination of Temporal Information and Linear Transformation of Feature Vector in Speaker Verification

서 창 우[1] · 조 미 화 · 임 영 환[2] · 전 성 채[3]
Seo, Changwoo · Zhao, Meihua · Lim, Younghwan· Jeon, Sungchae

ABSTRACT

The feature vectors which are used in conventional speaker recognition (SR) systems may have many correlations between their neighbors. To improve the performance of the SR, many researchers adopted linear transformation method like principal component analysis (PCA). In general, the linear transformation of the feature vectors is based on concatenated form of the static features and their dynamic features. However, the linear transformation which based on both the static features and their dynamic features is more complex than that based on the static features alone due to the high order of the features. To overcome these problems, we propose an efficient method that applies linear transformation and temporal information of the features to reduce complexity and improve the performance in speaker verification (SV). The proposed method first performs a linear transformation by PCA coefficients. The delta parameters for temporal information are then obtained from the transformed features. The proposed method only requires 1/4 in the size of the covariance matrix compared with adding the static and their dynamic features for PCA coefficients. Also, the delta parameters are extracted from the linearly transformed features after the reduction of dimension in the static features. Compared with the PCA and conventional methods in terms of equal error rate (EER) in SV, the proposed method shows better performance while requiring less storage space and complexity.

Keywords: speaker verification (SV), principal component analysis (PCA), delta cepstrum, Gaussian Mixture model (GMM).

1. 서 론

Gaussian Mixture model(GMM)을 기본으로 하는 화자인식(speaker recognition) 시스템에서 특징벡터를 추출할 때 관측된 특징벡터들이 상호 독립적으로 상관성(correlation) 없이 동일하게 분포되어 있다고 가정한다. 그러나 관측된 특징벡터들은 서로 상관성이 없지 않으므로 이 가정을 사용하는 경우 화자인식 시스템의 성능은 그만큼 저하될 수 밖에 없다[1,2]. 특히 음성의 정적 특징벡터(static feature vector)로부터 차분 켑스트럼(differential cepstrum)이나 델타 켑스트럼(delta cepstrum)과 같은 동적 특징벡터(dynamic feature vector)를 추출할 때, 회귀 방법(regression method)을 적용한 일반적인 방법에서는 특징벡터들간의 상관성에 직접적인 영향을 받고 있다.

최근에 이런 문제에 대한 해결방법으로 특징벡터들간의 상관성을 제거하고 차원을 감소시키기 위한 주성분분석(principal component analysis: PCA)이 널리 사용되고 있다[3,4]. 주성분분석은 입력된 음성데이터로부터 추출된 특징벡터들을 상관성이 없는 새로운 좌표계로 선형 변환(linear transformation)을 시킨다. 그러나 공분산 행렬(covariance matrix)에 의한 고유치(eigenvalue)와 고유벡터(eigenvector)를 구할 때, 정적인 특징벡터에 동적인 특징벡터를 추가할 경우, 주성분분석은 높은 차수의 특징벡터 때문에 많은 계산량이 필요하다[5,6].

본 논문에서는 이런 문제점을 해결하기 위해서 특징벡터에 순시 정보(temporal information)와 선형 변환 (linear transformation)을 효과적으로 적용하는 방법을 제안 하였다. 제안한 방법은 먼저 상관성이 높은 정적 특징벡터로부터 상관성을 줄이고 차원을 감소시키기 위해서 주성분분석을 적용하는 것이다. 다음으로, 상관성이 제거되고 차원이 축소된

1) 숭실대학교 cwseo@ssu.ac.kr, 교신저자
이 논문은 2009 년도 숭실대학교의 교내연구비 지원으로 이루어졌습니다.

접수일자: 2009 년 10 월 19 일
수정일자: 2009 년 11 월 18 일
게재결정: 2009 년 11 월 20 일

특징벡터로부터 순시 정보를 얻기 위한 동적인 특징벡터를 계산하는 것이다. 이런 과정으로 진행할 경우 세 가지의 잇점을 가질 수 있다. i)낮은 차수의 정적 특징벡터로부터 선형 변환을 수행하기 위한 주성분 분석에서 작은 공분산 행렬(covariance matrix)로부터 고유치와 고유벡터를 계산하기 때문에 계산량을 줄일 수 있다. ii)동적 특징벡터를 구할 때도 상관성이 없는 선형 변환된 특징벡터로부터 구할 수 있다. iii)동적 특징벡터를 계산할 때, 처음에 주어진 정적 특징벡터의 차수보다 낮은 특징벡터를 계산할 수 있다. 제안된 방법의 우수성을 확인하기 위해서 일반적인 방법과 주성분분석을 화자확인(speaker verification)에서 비교 실험의 결과로 설명하였다.

논문은 다음과 같이 구성되었다. 2 장에서는 특징벡터의 순시 정보를 설명하였다. 3 장에서는 특징벡터의 상관성 제거를 위한 선형 변환을 설명하고, 4 장에서는 제안한 특징벡터의 전체적인 구성을 기술하였다. 5 장에서는 Gaussian Mixture model (GMM)기반의 화자확인 시스템을 기술하였다. 그리고 6 장과 7 장에서는 실험 결과 및 결론을 서술하였다.

2. 특징벡터의 순시 정보(Temporal information)

음성인식과 화자인식 시스템의 성능은 기본적인 정적 특징벡터에 동적 특징벡터인 델타 켑스트럼(delta cepstrum) 을 추가함으로써 향상시킬 수 있다[7]. 먼저 프레임 길이가 T인 k-차 특징벡터 $x_k(t), t=1,\cdots,T$ 라 할 때, 정적 켑스트럼 계수 $x_k(t)$ 로부터 델타 켑스트럼은 다음과 같이 회귀 공식(regression formula)을 이용하여 계산할 수 있다.

$$\Delta x_k(t)=\frac{\sum_{\theta=1}^{\Theta}\theta(x_k(t+\theta)-x_k(t-\theta))}{2\sum_{\theta=1}^{\Theta}\theta^2} \tag{1}$$

여기서 $\Delta x_k(t)$ 는 대응하는 시간 t 의 정적계수 $x_k(t-\theta)$ 에서 $x_k(t+\theta)$ 까지 구간별 계산된 k-차 델타켑스트럼이고, Θ 의 값은 델타윈도우가 계산되는 파라미터의 시간 확장이다. 위 식은 과거와 미래의 음성 파라미터 값에 종속적이기 때문에, 음성의 시작과 끝부분에 대한 문제는 다음과 같이 1 차 차분을 이용해서 해결할 수 있다.

$$\Delta x_k(t)=x_k(t+1)-x_k(t),\ t<\Theta \tag{2}$$

$$\Delta x_k(t)=x_k(t)-x_k(t-1),\ t\geq T-\Theta \tag{3}$$

정적과 동적 특징벡터를 갖는 전체적인 특징벡터는 다음과 같이 $v(=k+k)$-차 특징벡터 Y_v 로 나타낼 수 있다.

$$Y_v=[x_k(t),\Delta x_k(t)] \tag{4}$$

3. 특징벡터의 선형 변환(Linear transform)

연속적인 특징벡터에서의 선형 변환(linear transformation)은 특징벡터들간의 신호의 상관성을 제거하고 효과적으로 모델링하기 위해서 사용된다. 이러한 선형 변환에서 널리 사용되는 방법으로는 주성분분석이 있다. 주성분 분석은 여러 개의 변수들에 대하여 얻어진 다변량 자료를 분석한 후 다차원적인 변수들을 축소, 요약함으로써 차원을 단순화시키고 서로 상관관계가 있는 반응 변수들간의 복잡한 구조를 분석하는데 목적이 있다[1,2]. 따라서 입력된 음성 데이터로부터 추출된 특징 벡터들을 상관관계가 없는 새로운 좌표계로 선형 변환시킨다. 이러한 특성 때문에 주성분분석은 특징벡터의 선형 변환방법에서 널리 사용된다.

데이터 집합에서 주성분을 찾는 일반적인 방법은 고유치 분할(eigenvalue decomposition) 방법을 이용한 공분산 행렬(covariance matrix)의 고유벡터(eigenvector)와 고유치(eigenvalue)를 계산하는 것이다[8]. 정적인 특징벡터에 동적인 델타 켑스트럼을 추가한 v-차 특징벡터 $Y=\{y_i(t),i=1,\cdots,v,t=1,\cdots,T\}$ 로부터 선형 변환 행렬 Ω^T 을 구하기 위해서 다음과 같이 전체 평균벡터(mean vector)와 공분산 행렬을 구해야 한다.

$$\mu_i=\frac{1}{T}\sum_{t=1}^{T}y_i(t) \qquad i=1,2,\cdots,v \tag{5}$$

$$\sigma_{ij}=\frac{1}{T}\sum_{t=1}^{T}(y_i(t)-\mu_i)^T(y_j(t)-\mu_j) \quad i,j=1,2,\cdots,v \tag{6}$$

$$\Sigma=\left[\begin{array}{cccc|ccc|c} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} & \sigma_{1k+1} & \cdots & \sigma_{1v-1} & \sigma_{1v} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} & \sigma_{2k+1} & \cdots & \sigma_{2v-1} & \sigma_{2v} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \iddots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} & \sigma_{kk+1} & \cdots & \sigma_{kv-1} & \sigma_{kv} \\ \hline \sigma_{k+11} & \sigma_{k+12} & \cdots & \sigma_{k+1k} & \sigma_{k+1k+1} & \cdots & \sigma_{k+1v-1} & \sigma_{k+1v} \\ \vdots & \vdots & \iddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{v-11} & \sigma_{v-12} & \cdots & \sigma_{v-1k} & \sigma_{v-1k+1} & \cdots & \sigma_{v-1v-1} & \sigma_{v-1v} \\ \sigma_{v1} & \sigma_{v2} & \cdots & \sigma_{vk} & \sigma_{vk+1} & \cdots & \sigma_{vv-1} & \sigma_{vv} \end{array}\right] \tag{7}$$

여기서 μ_i 는 i 번째 성분의 평균, Σ는 $v\times v$ 행렬로 σ_{ij} 를 원소로 갖는 공분산 행렬이다. Σ의 (i,j)번째 성분은 $i\neq j$ 일 때, Y의 i 번째와 j 번째 성분의 공분산을 나타내고 $i=j$ 일 때는 Y의 j 번째 성분의 분산을 나타낸다. 공분산 Σ는 다음과 같이 나타낼 수 있다.

$$\Sigma=\sum_{i=1}^{l}\lambda_i\omega_i\omega_i^T \tag{8}$$

여기서 λ_i 는 Σ의 i 번째 고유치(eigenvalue)이고, ω_i 는 고유값 λ_i 에 대응되는 정규화된 고유벡터(eigenvector) 이다. 이들은 $v\times v$ 인 직교 행렬(orthogonal matrix) $\Omega\Omega^T=I$ 을 이룬다. 위의 설명과 같이 t 번째 시퀀스의 i 번째 특징벡터 $y_i(t)$ 와 주성분 $z_i(t)$ 의 관계는

$$z_i(t)=v_i^T y_i(t)$$

(9)

이고, 성분 전체의 관계를 식으로 나타내면 다음과 같다.

$$Z = \Omega_p^T Y \tag{10}$$

식(10)에서 Ω_p^T는 v-차원 특징벡터 Y 를 차원 감소된 p-차원 주성분 Z 로 변환하기 위한 변환 행렬(transformation matrix)이다. v-차원 특징벡터에 대한 근사화의 의미인 p-차원 주성분 벡터의 정보율(information rate) α 은 다음 식에 의해 구할 수 있다.

$$\alpha = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{v} \lambda_i} \tag{11}$$

이 정보율에 따라 고유값이 큰 것부터 p-차원만을 선택하여 변환 행렬 Ω_p^T를 구하고, 식(10)과 같이 적용할 수 있다 [9].

4. 특징벡터의 구성도

본 장에서는 화자인식을 위한 정적 켑스트럼, 동적 켑스트럼, 그리고 주성분분석을 이용한 효과적인 특징 파라미터를 추출하는 방법을 설명하고자 한다. <그림 1>은 2 장과 3 장에서 설명한 순시 정보와 선형 변환의 일반적인 진행 과정을 나타낸 것이다[1,10,11]. 그림에서 정적 특징벡터로부터 동적 특징벡터를 구할 때, 첫 번째 타원(ellipse)에서 차수가 높은 k-차 정적 특징벡터로부터 동적 특징벡터를 계산하고 있다. 또한 두 번째 타원의 주성분분석은 식(4)에서 특징벡터가 높은 차수로 구성되기 때문에 식(7)과 같이 큰 공분산 행렬로부터 주성분 분석을 계산해야만 한다. 이 경우 $v \times v$ 크기의 큰 공분산 행렬로부터 고유치와 고유벡터를 얻기 위해서 많은 계산량을 요구한다.

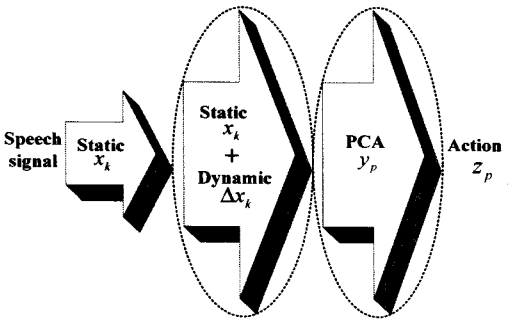

그림 1. 일반적인 특징벡터의 진행과정.
Figure 1. Progress process for the conventional feature vectors.

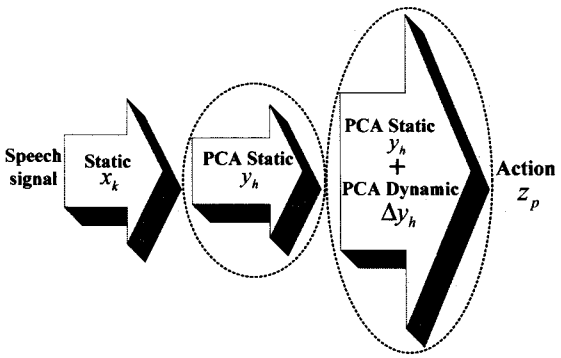

그림 2. 제안된 방법의 진행과정.
Figure 2. Progress process of the proposed method.

논문에서 제안한 특징벡터의 생성 과정은 <그림 2>와 같이 입력 음성으로부터 구해진 정적 특징벡터 x_k 를 구한 후 선형 변환과 순시 정보를 추출하였다. 이런 방법으로 진행할 경우, <그림 2>의 두 개의 타원으로부터 두 가지의 잇점을 갖는다. i)낮은 k-차수의 정적 특징벡터로부터 주성분분석을 수행하기 때문에 식(12)와 같이 공분산 행렬의 크기에서 기존 방법의 크기 $v \times v$ 보다 1/4 로 줄어든 $k \times k$ 행렬이 요구된다.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{11} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix} \tag{12}$$

ii)동적 켑스트럼 Δy_h 은 선형 변환된 켑스트럼 y_h 으로부터 구할 수 있다. iii)동적 켑스트럼을 계산할 때, 처음에 주어진 정적 특징벡터의 차수 k-보다 차원 축소된 $h(k \geq h = p/2)$ 만을 구하면 된다.

5. 화자확인 시스템

5.1 Gaussian Mixture model(GMM)

대각 공분산 행렬(diagonal covariance matrix)을 갖는 GMM 방법은 화자인식에서 가장 널리 사용되고 있다. 이런 가우시안 혼합 성분 밀도(Gaussian Mixture density)는 다음과 같이 M 성분 밀도함수의 가중화된 합으로 나타낼 수 있다[12].

$$p(z(t)|\theta) = \sum_{i=1}^{M} w_i b_i(z(t)), \tag{13}$$

여기서,

$$b_i(z(t)) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|^{1/2}} \exp\left\{ -\frac{1}{2} (z(t) - \mu_i)^T \Sigma_i^{-1} (z(t) - \mu_i) \right\}, \tag{14}$$

여기서 p 는 특징벡터의 차수이다. μ_i 는 평균 벡터(mean vector), Σ_i 은 공분산 행렬(covariance matrix)이다. 그리고 w_i 는 $\sum_{i=1}^{M} w_i = 1$ 조건을 만족시키는 M 차 혼합 성분을 위한 가중치(weight)를 나타낸다. 특징벡터 $Z = \{z(1), \cdots, z(T)\}$ 가 주어질 때, 화자 모델을 위한 GMM 은 모든 성분밀도로부터 평균 벡터 μ_i, 공분산 행렬 Σ_i, 그리고 가중치 벡터 w_i 는 다음과 같이 파라미터화 할 수 있다.

$$\theta = \{w_i, \mu_i, \Sigma_i\}_{i=1}^{M}. \tag{15}$$

이때, 파라미터 θ에 대한 GMM 의 유사도(likelihood)는 다음과 같이 나타낼 수 있다.

$$p(Z|\theta) = \prod_{t=1}^{T} p(z(t)|\theta) \tag{16}$$

일반적으로, GMM 의 유사도를 최대화시키는 파라미터 추정은 최대 유사도(maximum likelihood: ML) 방법이 널리 사용된다. 그러나, 비선형 함수의 GMM 유사도는 직접적으로 최대화를 얻을 수 없기 때문에 ML 추정은 기대치-최대화(expectation-maximization: EM) 알고리즘을 반복적으로 사용하여 구할 수 있다[13].

EM 알고리즘의 기본적인 진행은 초기 모델 θ로 시작해서, $p(Z|\hat{\theta}) \geq p(Z|\theta)$인 새로운 모델 $\hat{\theta}$를 추정하는 것이다. 새로운 모델은 다음 반복을 위해서 초기 모델이 되고 그리고 이러한 과정은 수렴 한계에 도달할 때까지 반복된다.

5.2 화자확인 시스템(Speaker verification system)

GMM 에서 화자 모델(speaker model)과 사칭자 배경 모델(universal background model: UBM)[14,15]을 위한 확률밀도함수(probability density function: PDF)를 $p(Z|\theta_s)$와 $p(Z|\theta_b)$라 하자. 화자확인에서 화자 모델과 사칭자 배경 모델을 이용한 평균 로그-유사도 비(log-likelihood ratio)는 다음과 같이 나타낼 수 있다.

$$\begin{aligned}\Lambda_m(Z|\theta_s,\theta_b) &= \frac{1}{T}\sum_{t=1}^{T}\log\frac{p(z(t)|\theta_s)}{p(z(t)|\theta_b)} \\ &= \frac{1}{T}\sum_{t=1}^{T}\log p(z(t)|\theta_s) - \frac{1}{T}\sum_{t=1}^{T}\log p(z(t)|\theta_b) \\ &= L(Z|\theta_s) - L(Z|\theta_b)\end{aligned} \tag{17}$$

이와 같이 유사도 비(likelihood ratio)를 이용한 결정 판별(decision threshold)은 다음과 같이 결정할 수 있다.

$$\Lambda_m(Z|\theta_s,\theta_b) \begin{cases} \geq Th, & accepts\ speaker \\ < Th, & rejects\ imposter \end{cases} \tag{18}$$

여기서 Th는 결정 로직(decision logic)을 위한 판별 값이다. 지금까지 설명된 전체적인 화자확인 시스템의 구성도는 <그림 3>과 같이 나타낼 수 있다.

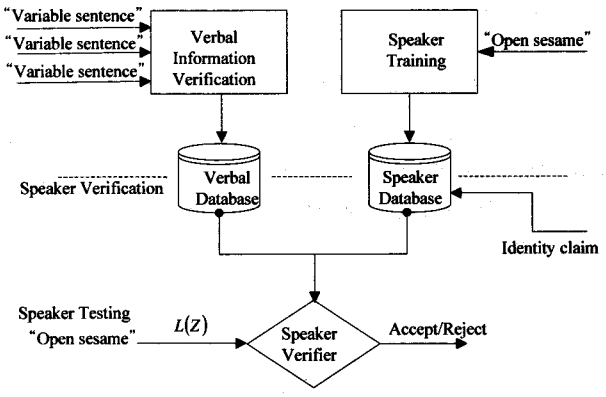

그림 3. 화자확인 시스템의 구성도.

Figure 3. An overview of speaker verification system.

6. 실험 결과

제안된 알고리즘의 성능을 검증하기 위해서 제안된 방법(Static → PCA + Dynamic), 일반적인 특징벡터(Static + Dynamic), 그리고 주성분분석 방법(Static + Dynamic → PCA)을 화자확인 방법에서 비교 실험하였다. 수집한 데이터는 발성길이가 평균 0.8 초로 짧은 문장인 "열려라 참깨(Open sesame)"와 발성길이가 평균 2.0 초로 비교적 긴 문장인 "무궁화 꽃이 피었습니다(Roses of Sharon have blossomed)"를 사용하였다. 실험에 사용된 음성 데이터는 200 명의 화자(남·여 각각 100 명)로부터 획득하였고, 개인별 화자의 데이터는 주 단위로 3 주 동안 15 개(주 5 문장)의 데이터를 구하였다. 처음 2 주 동안 수집한 10 개의 데이터를 학습(training)에 사용

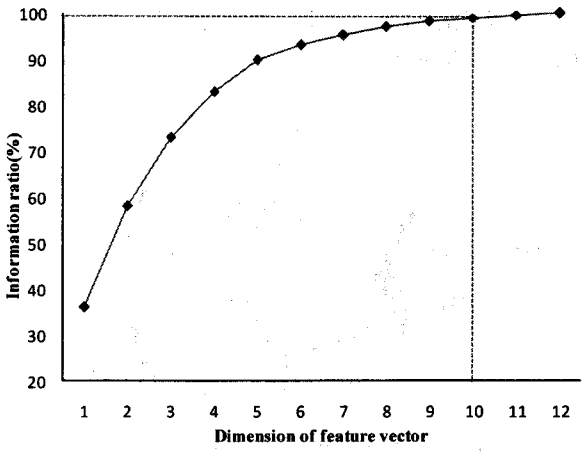

그림 4. 누적된 고유치 계수의 정보율.

Figure 4. Information ratio of the accumulated eigenvalue coefficients.

하였고, 마지막 주에 수집한 나머지 5 개를 확인(verification) 과정에 사용하였다. 따라서 테스트에서 사용된 오거절율(false reject rate: FRR)과 오인증율(false accept rate: FAR)을 위한 문장은 각각 1,000(200×5)와 199,000 (200×199×5)이다.

음성의 분석과정에서 샘플링 주파수는 11.025kHz, 16bit 분해능, 그리고 음성분석 프레임은 16ms 이고 50% 중첩을 적용하였다. 정적 특징벡터를 위한 12 차 mel frequency cepstral coefficient(MFCC) 그리고 동적 특징벡터를 위한 12 차 델타 켑스트럼(delta cepstrum)을 사용하였다. 동적 특징벡터를 위한 델타 켑스트럼의 Θ는 2 를 사용하였다. 선형 변환에서의 차원 감소는 <그림 4>와 같이 고유치의 정보율을 99%까지 포함하는 10 차를 기준으로 하였다. 사칭자 배경 모델(UBM)을 위한 데이터는 학습과 테스트에 참여하지 않은 50 명의 화자로부터 발성된 개인별 10 개의 문장-독립(text-independent) 데이터 500×(50×10)를 이용하여 모델을 구성하였다. 사칭자 배경 모델은 250 개의 혼합성분 (mixture)으로 구성되었고, 초기치 추정은 vector quantization(VQ) 클러스터링을 이용하였다[16].

표 1. 제안된 방법, PCA, 그리고 일반적인 방법의 파라미터 수.
Table 1. Required number of parameters for the proposed, PCA, and conventional methods.

Proposed method	PCA	Conventi. method
$M(2p+1)+k\times h$	$M(2v+1)+v\times v$	$M(2v+1)$

k: original feature vector (static: 12)
h: reduced feature vector (PCA static: 10), $h \le k$
v: extended feature vector (static + dynamic: 24), $v \ge p$
p: extended feature vector (PCA static + PCA dynamic: 20)
M: number of Mixtures (24, 32, 48)

<표 1>은 실험에서 사용된 제안된 방법, 일반적인 방법, 그리고 PCA 방법에서 요구되는 식(15)의 파라미터 수를 나타낸 것이다. 비록 제안된 방법에서 선형 변환을 위한 고유벡터의 크기가 $k\times h$ 개를 요구하지만, 이것은 $v\times v$ 개 를 요구하는 PCA 방법보다 공분산 행렬의 크기를 1/4 로 줄일 수 있다. 그리고 M = 32 에서 전체 파라미터를 비교 했을 때, 일반적인 방법과 PCA 방법이 각각 32×(2×24+1) = 1568 와 32×(2×24+1)+24×24 = 2144 를 요구하지만, 제안된 방법은 32×(2×20+1)+12×10 = 1432 개로 약 8.7%와 33% 정도 감소하였다.

<표 2>는 음성 데이터 "무궁화 꽃이 피었습니다(Roses of Sharon have blossomed)"를 GMM 의 다양한 혼합성분 수에서 FRR, FAR, 그리고 등가오율(equal error rate: EER)을 변화시키면서 화자확인을 수행한 결과이다. 실험 결과로부터 같은 차수의 특징벡터와 혼합성분을 사용할 때, EER 에서 제안된 방법(24)을 일반적인 방법(24)과 PCA 방법(24)을 비교했을 때 평균 0.45%와 0.13% 높은 확인 성능을 보였다. 또한 차원이 감소된 제안한 방법(20)은 일반적인 방법(24)보다 파라미터 수에서 약 8.3% 작지만 약간 우수한 성능을 보였고, 차원감소에 의한 PCA 방법(24)보다는 약간 감소하였다. 그리고 PCA(24)를 일반적인 방법(24)과 비교했을 때, 계산량은 비록 증가되었지만 상관성 제거에 의한 성능이 평균 0.32% 향상되었다.

표 2. 다양한 혼합성분에서의 화자확인 오류(%).
Table 2. Speaker verification errors (%) for various mixtures.

	24			32			48		
	FRR	FAR	EER	FRR	FAR	EER	FRR	FAR	EER
Conventional method(24)	7.1	1.17	**3.42**	7.1	1.12	**3.27**	6.9	1.12	**3.08**
PCA(24)	7.0	1.12	**3.05**	6.9	1.11	**2.96**	7.0	1.03	**2.83**
Proposed method(24)	6.9	1.08	**2.93**	6.9	0.98	**2.82**	6.9	0.93	**2.67**
Proposed method(20)	7.2	1.13	**3.34**	7.1	1.12	**3.18**	7.1	1.06	**2.97**

<표 3>은 비교적 발성길이가 짧은 "열려라 참깨(Open sesame)"를 혼합성분의 개수를 변화시키면서 나타낸 것이다. 실험 결과로부터 학습과 테스트 데이터가 충분하지 못할 경우 높은 차수의 특징벡터에 선형 변환을 이용한 PCA 방법(24)은 일반적인 방법(24)과 거의 비슷한 성능이 나왔다. 그렇지만 낮은 차수에서 선형 변환을 수행하고 상관성이 제거된 상태에서 동적 특징벡터를 구한 제안한 방법(24)에서는 일반적인 방법(24)과 PCA 방법(24)보다 평균 0.65%와 0.52% 향상되었다. 그리고 차원이 감소된 제안한 방법(20)을 일반적인 방법(24)과 PCA 방법(24)에 비교했을 때, 평균 0.19%와 0.06% 높은 성능 결과를 보였다. 제안된 방법이 기존의 방법들보다 우수한 결과를 보인 것은 순시 정보를 추출할 때, 상관성이 없는 선형 변환된 특징벡터로부터 획득하였기 때문이다.

따라서 제안된 방법은 화자인식의 특성상 계산량 감소와 특징벡터의 상관성 제거에 의한 성능향상에 효과적인 방법이라 할 수 있다.

7. 결 론

본 논문에서는 화자인식(speaker recognition)에서 널리 사용되고 있는 특징벡터의 순시 정보(temporal information)와 선형 변환(linear transformation)을 효과적으로 이용하여 계산량을 줄이고 성능을 향상시키기 위한 방법을 제안하였다. 제안된 방법은 주성분분석(principal component analysis: PCA)을 이용한

선형 변환을 수행하고 그리고 순시 정보를 위한 델타 켑스트럼(delta cepstrum)을 추출하는 방법을 적용하였다. 제안한 방법과 같이 진행할 경우, 주성분분석을 위한 공분산 행렬(covariance matrix)의 크기를 1/4 로 줄일 수 있을 뿐 아니라, 순시 정보를 추출할 때 상관성이 없는 선형 변환된 특징벡터로부터 얻을 수 있었다. 제안된 방법의 우수성을 확인하기 위하여 화자확인에서 일반적인 방법과 PCA 방법을 비교했을 때, 작은 계산량으로 높은 성능을 보였다.

표 3. 다양한 혼합성분에서의 화자확인 오류(%).

Table 3. Speaker verification errors (%) for various mixtures.

	24			32			48		
	FRR	FAR	EER	FRR	FAR	EER	FRR	FAR	EER
Conventional method(24)	8.0	1.21	**3.95**	7.7	1.12	**3.75**	7.8	1.10	**3.69**
PCA(24)	7.7	1.28	**3.83**	7.6	1.10	**3.67**	7.7	1.08	**3.51**
Proposed method(24)	7.7	1.06	**3.34**	7.7	1.03	**3.16**	7.8	0.98	**2.95**
Proposed method(20)	7.6	1.13	**3.76**	7.8	1.09	**3.61**	7.6	1.07	**3.47**

참 고 문 헌

L. Liu, and J. He, "On the use of orthogonal GMM in speaker recognition", *ICASSP 99*, pp. 845-849, 1999.

Y. Ariki, S. Tagashira, and M. Nishijima, "Speaker recognition and speaker normalization by projection to speaker subspace", *ICASSP 96*, pp. 319-322, 1996.

C. Seo, and Y. Lim, "Global covariance based principal component analysis for speaker identification", *Jour. of Kor. Soc. Spee. Sci.*, Vol. 1, No. 1, pp. 69-73, 2009.
(서창우, 임영환, "화자식별을 위한 전역 공분산에 기반한 주성분분석", 말소리와 음성과학, Vol. 1, No. 1, pp. 69-73, 2009.

T. Takiguchi and Y. Ariki, "Robust Feature Extraction using Kernel PCA", *ICASSP 06*, pp. 509-512, 2006.

N. Kambhatla, and T.K. Leen, "Dimension reduction by local PCA", *Neural Computation*, Vol. 9, pp. 1493-1503, 1997.

C. Seo, K.Y. Lee, and J.H. Lee, "GMM based on local PCA for speaker identification", *Electronic Letters*, Vol. 37, No. 24, pp. 1486-1488, Nov., 2001.

S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (for HTK version 3.2)*, Cambridge University Engineering Department, 2002.

G.H. Golub, and C.F.V Loan, *Matrix Computations*, The Johns Hopkins University Press, third ed., 1996.

B.N. Flury, "Common Principal Components in k Groups," *JASA*, Vol. 79, No. 388, pp. 892-898, Dec, 1984.

O. Thyes, R. Kuhn, P. Nguyen and J-C. Junqua, "Speaker Identification and Verification Using Eigenvoices", *ICSLP-2000*, Vol. 2, pp. 242-245, Oct. 2000.

K.Y. Lee, "Local fuzzy PCA based GMM with dimension reduction on speaker identification", *Pattern Recog. Letters*, Vol. 25, pp. 1811-1817, 2004.

D. Reynolds, and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models" , *IEEE Trans. on SAP*, Vol. 3, No. 1, pp. 72-82, 1995.

A. Dempster, N. Laird, and D. Doubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Royal Stat. Sco.*, Vol. 29, pp. 1-38, 1977.

S. Parthasarathy and A.E. Rosenberg, "General phrase speaker verification using sub-word background models and likelihood-ratio scoring", *ICSLP-96*, 1996.

T. Matsui and S. Furui, "Likelihood normalization for speaker verification using a phoneme and speaker-independent model", *Speech Communication*, Vol. 17, pp. 109-116, August, 1995.

Y. Lind, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design", *IEEE Trans. Commun.*, Vol. 28, pp. 84-95, 1980.

• **서창우 (Seo, Changwoo) 교신저자**
숭실대학교 글로벌 미디어학부
서울시 동작구 상도동 511번지
Tel: 02-826-9872 Fax: 02-826-9872
Email: cwseo@ssu.ac.kr
관심분야: 음성신호처리, 멀티미디어, 모바일 시스템
2008~현재 글로벌 미디어학부 연구교수

• **조미화 (Zhao, Meihua)**
숭실대학교 글로벌 미디어 학부
서울시 동작구 상도동 511번지
Email: meehwa@ssu.ac.kr
관심분야: 멀티미디어, 모바일 시스템
2008~현재 미디어학과 박사과정

• **임영환 (Lim, Younghwan)**
숭실대학교 글로벌 미디어학부
서울시 동작구 상도동 511번지
Tel: 02-820-0685 Fax: 02-820-0685
Email: yhlim@ssu.ac.kr
관심분야: 멀티미디어, 모바일 시스템
1996~현재 글로벌 미디어학부 교수

• **전성채 (Jeon, Sungchae)**
한국전기연구원 융합기술연구단
경기도 안산시 상록구 사1동 1271-19번지
Tel: 031-8040-4151 Fax: 031-8040-4163
Email: sarim@keri.re.kr
관심분야: 신호처리, 방사선 센서, 방사선 신호검출용 ASIC 설계
2006~현재 전자의료기기 연구센터 선임연구원