

커널 주성분 분석의 앙상블을 이용한 다양한 환경에서의 화자 식별

Speaker Identification on Various Environments Using an Ensemble of Kernel Principal Component Analysis

양일호 · 김민석* · 소병민 · 김명재 · 유하진

(Il-Ho Yang, Min-Seok Kim*, Byung-Min So, Myung-Jae Kim, and Ha-Jin Yu)

서울시립대학교 컴퓨터과학부, *LG 전자기술원

(접수일자: 2011년 12월 8일; 수정일자: 2012년 1월 17일; 채택일자: 2012년 3월 21일)

초 록: 본 논문에서는 커널 주성분 분석(KPCA, kernel principal component analysis)으로 강화한 화자 특징을 이용하여 복수의 분류기를 학습하고 이를 앙상블 결합하는 화자 식별 방법을 제안한다. 이 때, 계산량과 메모리 요구량을 줄이기 위해 전체 화자 특징 벡터 중 일부를 랜덤 선택하여 커널 주성분 분석의 기저를 추정한다. 실험 결과, 제안한 방법이 그리디 커널 주성분 분석(GKPCA, greedy kernel principal component analysis)보다 높은 화자 식별률을 보였다.

핵심용어: 화자식별, 분류기 앙상블, 주성분 분석, 커널 주성분 분석, 그리디 커널 주성분 분석

투고분야: 음성처리 분야(2.5)

ABSTRACT: In this paper, we propose a new approach to speaker identification technique which uses an ensemble of multiple classifiers (speaker identifiers). KPCA (kernel principal component analysis) enhances features for each classifier. To reduce the processing time and memory requirements, we select limited number of samples randomly which are used as estimation set for each KPCA basis. The experimental result shows that the proposed approach gives a higher identification accuracy than GKPCA (greedy kernel principal component analysis).

Key words: Speaker identification, Classifier ensemble, PCA, KPCA, GKPCA

ASK subject classification: Speech Signal Processing (2.5)

1. 서 론

화자 인식 시스템의 정확도는 채널 특성이나 주변 잡음과 같은 인식 환경에 따라 하락할 수 있다. 이러한 문제를 완화하기 위해 MFCC (Mel frequency cepstral coefficient)와 같은 기존의 화자 특징을 환경 변화에 강인한 특징으로 변환하는 다양한 특징 강화 방법이 연구되어 왔다.

주성분 분석 (PCA, principal component analysis)^[1]은 널리 쓰이는 특징 강화 방법의 하나이다. 그러나

이는 비선형으로 분포한 데이터를 적절히 처리할 수 없다는 단점이 있다. 반면에 커널 주성분 분석 (KPCA, kernel principal component analysis)^[2,3]은 비선형으로 분포한 데이터를 다룰 있으나, 계산량과 메모리 요구량이 특징 벡터 수의 제곱에 비례하여 급격히 증가한다. 일반적으로 화자 인식 분야에서는 불과 몇 초 분량의 발성에서도 수백 개의 특징을 추출하므로 커널 주성분 분석을 그대로 적용하기는 어렵다. 이러한 문제를 해결하기 위해, 그리디 커널 주성분 분석 (GKPCA, greedy kernel principal component analysis)^[4]은 그리디 필터링 (greedy filtering)을 통해 특징 벡터의 수를 줄이는 전략을 취한다. 그리디 필

*Corresponding author: 교신저자누락 (이메일 누락)
주소누락
(전화: 누락)

터링은 전체 특징 벡터를 잘 대표하는 소수의 부분 특징 벡터를 선택한다.

본 연구에서는 분류기 앙상블^[5]의 개념을 도입함으로써 그리디 커널 주성분 분석의 한계를 극복하고자 했다. 제안한 방법에서는 그리디 커널 주성분 분석과 유사하게 전체 특징 벡터의 일부만을 사용하여 커널 주성분 분석을 수행한다. 이 때, 그리디 필터링을 사용하는 것이 아니라 랜덤하게 부분 특징 벡터를 선택한다. 이와 같은 과정을 여러 번 반복하여 복수 개의 부분 특징 벡터 집합을 얻은 후 각각을 이용하여 서로 다른 커널 주성분 분석의 기저를 추정한다. 이 기저들로 사상한 특징 벡터를 이용하여 복수 개의 분류기 (화자 식별기)를 학습하고 앙상블을 구성한다. 최종적인 분류기 앙상블의 결과는 다수 투표 (majority voting)^[5] 방식으로 도출한다.

본 논문의 구성은 다음과 같다. 2장과 3장에서 관련 연구와 제안한 앙상블 시스템을 소개한다. 4장에서 실험 결과를 보이고, 5장에서 결론을 맺는다.

II. 관련 연구

2.1 GMM-UBM을 이용한 화자 식별

GMM-UBM^[6]은 가우시안 혼합 모델 (GMM, Gaussian mixture model)^[7]로 배경 화자 모델 (UBM, universal background model)을 구축하고 각 화자의 학습 발성으로 MAP (maximum a posteriori) 적용을 수행함으로써 화자 모델을 학습한다.

가우시안 혼합 모델 (GMM, Gaussian mixture model)^[7]은 화자 인식 분야에서 널리 쓰이는 모델로서 여러 개의 가우시안 확률 밀도 함수를 결합한 형태로 표현된다. m 번째 혼합 성분의 결합 가중치와 가우시안 확률 밀도 함수를 각각 $w_m, p_m(\cdot)$ 라 할 때, 모델 파라미터 λ 를 지닌 혼합 수 M 개 가우시안 혼합 모델의 우도 (likelihood) 함수는 식 (1)과 같다.

$$p(\vec{x}|\lambda) = \sum_{m=1}^M w_m p_m(\vec{x}) \tag{1}$$

이 때, 입력 \vec{x} 는 D 차원의 특징 벡터이다. m 번째 혼합 성분의 가우시안 확률 밀도 함수 $p_m(\cdot)$ 는 D 차원

평균 벡터 $\vec{\mu}_m$ 와 $D \times D$ 크기의 공분산 행렬 Σ_m 에 의해 식 (2)와 같이 표현된다.

$$p_m(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_m)^T (\Sigma_m)^{-1} (\vec{x} - \vec{\mu}_m)\right\} \tag{2}$$

GMM-UBM에서 배경 화자 모델은 하나의 가우시안 혼합 모델로 표현되며 반복적인 EM (expectation-maximization) 알고리즘을 통해 학습된다. 이 때, 배경 화자 모델은 일반적인 사람의 음성을 표현해야 하므로 많은 화자의 다양한 발성으로 학습을 수행한다. 이렇게 학습한 배경 화자 모델에 각 화자의 학습 발성을 MAP 적용하여 최종적인 화자 모델을 얻는다.

인식 단계에서는 식 (3)과 같이 주어진 발성의 특징 벡터열 $\vec{X} = \{\vec{x}_1, \dots, \vec{x}_N\}$ 에 대한 로그 우도를 각 화자 모델에 대해 계산한다.

$$\log p(\vec{X}|\lambda) = \sum_{n=1}^N \log p(\vec{x}_n|\lambda) \tag{3}$$

2.2 주성분 분석

주성분 분석 (PCA, principal component analysis)^[1]은 널리 쓰이는 특징 강화 방법의 하나로서, 전체 데이터의 분산을 최대화하는 축으로 기존의 특징을 사상한다. 그림 1은 주성분 분석을 이용하여 2차원 특징을 1차원으로 사상하는 예를 보여준다. 여기서 원과 직선은 각각 특징 벡터와 사상 축을 나타낸다. 이 때, 왼쪽과 같이 특징이 선형으로 분포해 있을 경우에는 주성분 분석이 적합한 사상 축을 찾지만 오른쪽과 같이 비선형으로 분포해 있을 경우에는 적합한 사상 축을 찾지 못할 수 있다.

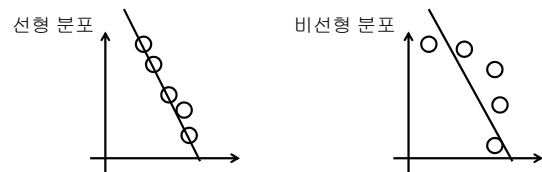


그림 1. 주성분 분석의 예
Fig. 1. examples of PCA.

2.3 커널 주성분 분석

커널 주성분 분석 (KPCA, kernel principal component analysis)^[2,3]은 비선형 분포를 처리할 수 있는 주성분 분석이다. 그림 2는 입력 공간 (input space)상에서 비선형으로 분포한 특징을 선형 분리 가능하게 하는 고차원 특징 공간 (feature space)상으로 사상하여 주성분 분석을 수행하는 예를 보여준다. 이 때, 특징을 고차원으로 사상하는 매핑 함수 $\phi(\cdot)$ 를 직접 구하기는 어려우나 특징 공간에서의 내적을 직접 구하는 커널 대치를 이용하여 문제를 해결할 수 있다. 그러나 커널 방법은 계산량과 메모리 요구량이 특징 벡터 수의 제곱에 비례하여 급격히 증가하므로 화자 인식 분야에 그대로 적용하기는 어렵다.

2.4 그리디 커널 주성분 분석

그리디 커널 주성분 분석 (GKPCA, greedy kernel principal component analysis)^[4]은 그림 3과 같이 그리

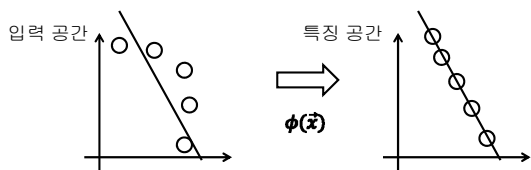


그림 2. 커널 주성분 분석의 예
Fig. 2. example of KPCA.

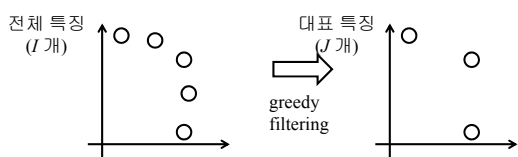


그림 3. 그리디 필터링의 예
Fig. 3. example of greedy filtering.

디 필터링 (greedy filtering)을 통해 I 개의 전체 특징 벡터 중에서 J 개의 부분 특징 벡터를 선택하고 ($I \ll J$), 이를 이용하여 커널 주성분 분석을 수행한다. 커널 주성분 분석 수행시 사용되는 특징 벡터의 수를 절감할 수 있으므로 계산량 및 메모리 요구량이 감소한다. 그림 4는 그리디 커널 주성분 분석을 이용한 인식기의 개략적인 수행 과정을 나타낸 것이다.

이를 GMM-UBM 방법을 이용하는 화자 인식 시스템을 기준으로 상세히 도식화하면 그림 5와 같이 표현할 수 있다.

GMM-UBM 방법에서 그리디 커널 주성분 분석을 수행할 때에는 전체 UBM 학습 데이터로부터 그리디 필터링을 통해 부분 학습 특징을 선택하고, 이로부터 추정된 커널 주성분 분석 변환 기저로 모든 특징 벡터를 사상한다. 하지만 화자 인식 분야에서는 일반적으로 전체 특징 벡터의 수 I 가 매우 크고 J 가 매우 작기 때문에, 비록 이 부분 특징 벡터가 질적인 면에서 전체 특징 벡터를 잘 대표하더라도 양적인 면에서는 충분하지 않을 수 있다. 제안한 방법은 이 한계를 극복하기 위해 여러 개의 인식기를 결합한 앙상블 시스템이다.

III. 제안한 방법

본 연구에서는 커널 주성분 분석을 이용한 앙상블 시스템을 제안한다. 제안한 방법은 전체 특징 벡터 중 일부를 랜덤하게 선택하여 커널 주성분 분석의 기저를 찾고, 이 기저로 사상한 특징을 사용해 화자 모델을 학습한다. 즉, 그림 5 상의 그리디 필터링을 랜덤 선택으로 대체한 것이다. 전체 데이터를 잘 표현하는 그리디 필터링 결과 대신 랜덤 선택 방식을

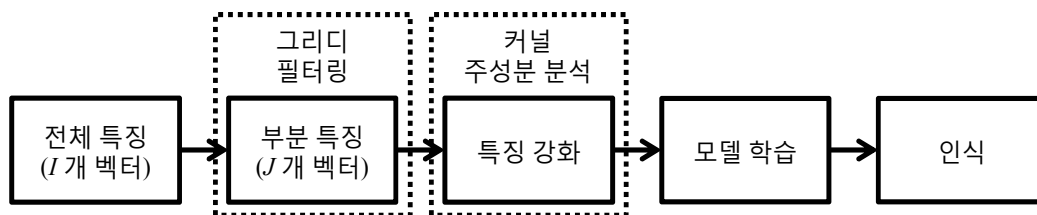


그림 4. 그리디 커널 주성분 분석을 이용한 인식기의 수행 과정
Fig. 4. process of recognition system using GKPCA.

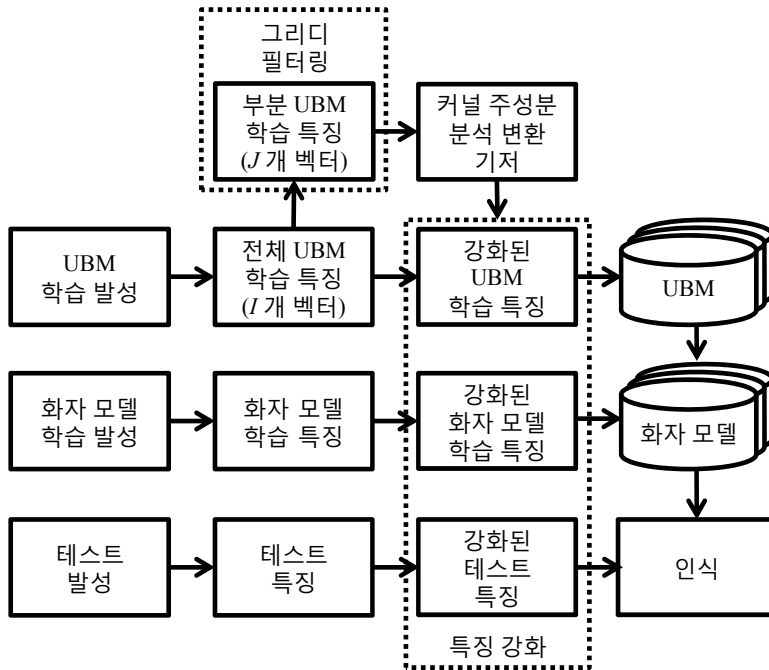


그림 5. GMM-UBM 방법에서의 그리디 커널 주성분 분석 수행 과정
 Fig. 5. process of GKPCA with GMM-UBM approach.

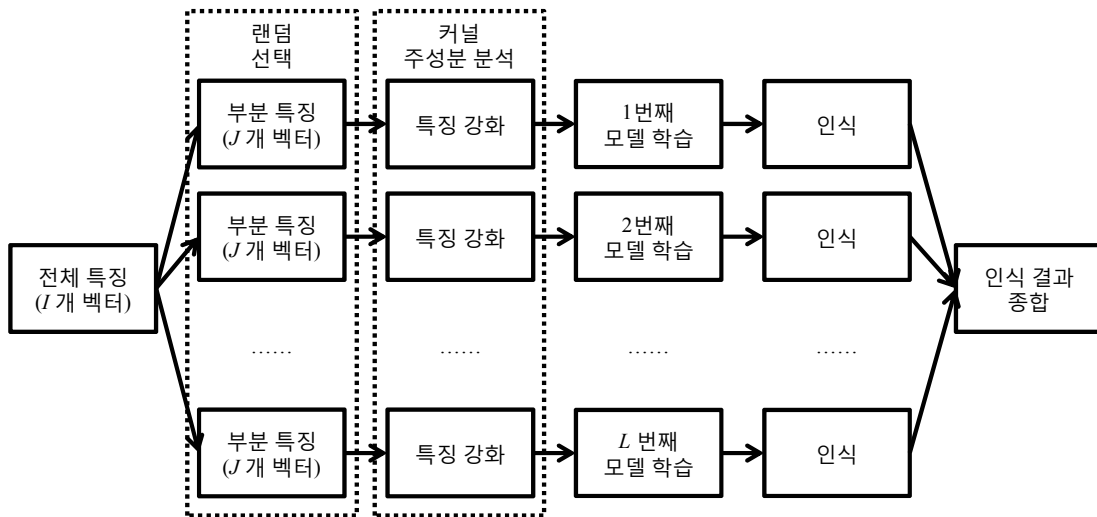


그림 6. 제안한 방법의 수행 과정 (m개 분류기 결합)
 Fig. 6. process of proposed method (combination of m classifiers).

취할 경우 물론 단일 분류기로서의 인식 정확도는 하락할 것이다. 대신 동일한 데이터에 대한 그리디 필터링 결과는 항상 동일하지만 랜덤 선택 결과는 항상 다르다. 본 연구에서는 이러한 랜덤 선택 과정을 여러 번 반복하여 얻은 복수 개의 분류기를 앙상

블 결합하여 화자 식별을 수행한다. 그림 6은 L개 분류기를 결합한 제안한 방법의 수행 과정을 나타낸 것이다. 이 때, 각 부분 특징에 대한 모델 학습 단계의 결과는 모든 화자를 다중 클래스 분류하는 독립적인 분류기이다.

제안한 방법은 기저 추정을 위한 부분 특징을 랜덤하게 선택하므로 그리디 필터링을 통해 전체 분포를 잘 대표하는 특징을 선택하는 그리디 커널 주성분 분석에 비해 식별률이 낮을 것으로 기대한다(단일 분류기를 사용할 경우). 하지만 분류기 앙상블의 개념을 도입하여 이러한 분류기를 복수 개 결합함으로써 보다 많은 특징 벡터를 사용하고자 하였다.

제안한 방법은 양적인 면에서 볼 때 기존의 그리디 커널 주성분 분석이 기저 추정에 이용하는 J 개의 특징 벡터보다 L 배 많은 $J \times L$ 개의 화자 특징 벡터를 사용한다. 기저 추정에 드는 계산량은 L 에 비례하여 선형적으로 증가하므로 J 개의 화자 특징을 사용하여 그리디 커널 주성분 분석을 수행하는 것보다는 많으나 $J \times L$ 개의 화자 특징을 모두 사용하여 커널 주성분 분석을 수행하는 것보다는 적다. 또한 앙상블을 구성하는 각각의 분류기를 순차적으로 학습할 수 있으므로 메모리 요구량은 그리디 커널 주성분 분석과 동일하다. 그러므로 사용 가능한 메모리가 한정된(커널 주성분 분석의 적용이 불가능한) 상황에서 제안한 방법이 그리디 커널 주성분 분석보다 더 많은 화자 특징을 이용할 수 있기 때문에 보다 높은 식별률을 보일 것으로 기대할 수 있다.

IV. 실험 설계 및 결과 분석

4.1 사용한 데이터베이스 및 화자 모델 학습

다양한 환경에서 제안한 시스템의 성능을 평가하기 위해 ETRI 중가마이크 화자인식용 DB(이하 'PC DB')와 ETRI 화자인식용 휴대전화 DB(이하 '휴대전화 DB')를 실험에 사용하였다. 이 코퍼스들은 발성 시차의 종류에 따라 주차, 월차, 3개월차로 화자가 나누어져 있다. 본 연구에서는 배경 화자 모델(UBM, universal background model) 학습에는 월차 화자의 10개 문장 발성(1월차 1회차 발성)을 이용하였고, 화자 모델 학습 및 테스트에는 주차 화자의 10개 문장 발성(학습: 1주차 1회차 발성, 테스트: 3주차 1회차 발성)을 이용하였다. PC DB 실험시에는 배경 화자 모델 학습, 화자 모델 학습 및 테스트에 각각 100명분의 발성(문장 10개씩 총 1000개 발성)을 사용하였고, 휴대전화 DB 실험시에는 배경 화자 모델 학습에 101명

(총 1010개 발성), 화자 모델 학습 및 테스트에 104명분의 발성(총 1040개 발성)을 사용하였다.

화자 모델 학습에는 GMM-UBM^[6] 방법을 사용하였다. 혼합 수 64개와 256개의 가우시안 혼합 모델(GMM, Gaussian mixture model)^[7]로 각각 배경 화자 모델을 구성하여 실험하였다. 가우시안 혼합 모델 학습시 혼합 수는 1개로부터 시작하여 2배씩 늘려갔다. 이 때, 각 혼합 수에서 모델 파라미터는 1회씩 학습하였고, 마지막 혼합 수(64개 및 256개)에서는 10회 반복 학습하였다. 화자 모델은 배경 화자 모델로부터 1회 MAP 적응^[6]하였다($\tau=1$).

제안한 시스템을 잡음 환경에서도 평가하기 위해 동일한 발성에 Aurora2 DB^[8]의 CAR, SUBWAY, RESTAURANT 잡음을 각각 20 dB와 10 dB로 삽입하였다. 잡음 삽입에는 FaNT^[9]를 사용하였다.

4.2 특징 추출

15차 MFCCs와 에너지, 그리고 이의 delta를 포함하여 총 32차 특징을 추출하였다(window size = 25 ms, shift = 10 ms). 무음은 특징 추출 후 에너지 기반으로 제거하였다. 추출한 특징에 대해 CMVN(cepstral mean and variance normalization)을 적용하였다.

4.3 특징 강화

배경 화자 모델 학습 특징으로부터 특징 강화를 위한 변환 기저를 추정된 뒤 모든 특징(배경 화자 모델 학습, 화자 모델 학습 및 테스트 특징)을 사상하였다. 사상한 특징의 차원은 원 특징과 동일한 32차이다. 제안한 방법과의 식별률 비교를 위해 주성분 분석과 그리디 커널 주성분 분석을 이용하였다.

커널 주성분 분석을 수행할 때에는 식 (4)와 같은 가우시안 커널 함수를 사용하였다($\sigma=32$).

$$k(\vec{x}, \vec{y}) = \exp\left(-\frac{\|\vec{x} - \vec{y}\|^2}{2\sigma^2}\right) \quad (4)$$

그리디 커널 주성분 분석에서의 그리디 필터링과 제안한 방법에서의 랜덤 선택을 수행할 때에는 배경 화자 모델 학습 특징으로부터 100개($J=100$)의 부분 특징을 선택하였다. 제안한 앙상블 시스템은 100개

표 1. 변환 기저 추정에 이용한 특징 수

Table 1. number of feature vectors which used as estimation set of transform basis.

항목	표기	해당 특징 강화 방법	DB	
			PC	휴대전화
전체 특징 수	t	PCA	168896개	191071개
선택한 부분 특징 수	n	GKPCA	100개	
양상블 결합시 최대 이용 가능 특징 수	$n \times m$	proposed	10000개	

의 서로 다른 분류기로 구성 ($L = 100$)하였다. 표 1은 변환 기저 추정에 이용한 특징 수를 나타낸다. 표에서 ‘PCA’, ‘GKPCA’, ‘proposed’는 각각 주성분 분석, 그리디 커널 주성분 분석, 제안한 방법을 의미한다. 제안한 방법에서의 랜덤 선택시 중복을 허용하였으므로 양상블 결합시 실제 이용 특징 수는 $n \times m$ 보다 작을 수 있다.

4.4 분류기 결합 방법

본 연구에서는 분류기 양상블의 결합 방법으로 다수 투표 (majority voting)와 Borda 계수를 사용하였다. 다수 투표 방법은 가장 많은 분류기로부터 식별된 화자를 양상블 시스템의 최종 식별 결과로 취하는 것이다. C 명의 전체 화자 (부류) 집합 $S = \{S_1, S_2, \dots, S_C\}$ 중에서 식별 화자 S_q 를 결정하는 다수 투표 방법은 다음과 같다.

$$S_q = \underset{S_c \in S}{\operatorname{argmax}} \left(\sum_{l=1}^L d_{l,S_c} \right) \quad (5)$$

$$d_{l,S_c} = \begin{cases} 1 & \text{if } l\text{번째 분류기의 식별 결과가 화자 } S_c \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

여기서 d_{l,S_c} 는 화자 S_c 에 대한 l 번째 분류기의 식별 결과 (0 혹은 1)를 의미한다. 식 (6)과 같이 만약 l 번째 분류기의 식별 결과가 S_c 였다면 d_{l,S_c} 는 1 그렇지 않은 경우 0이 된다.

Borda 계수는 전체 C 명 화자에 대한 한 분류기의 인식 결과를 확률이 큰 순서대로 $C-1$ 점에서 0점까지 부여하는 투표 방식이다. 본 연구에서는 한 분류기의 인식 결과에서 로그 우도 (likelihood)가 큰 순으로

5명 화자에 대해 5점에서 1점까지 부여하였다.

4.5 실험 결과

표 2는 두 데이터베이스 (PC, 휴대전화DB)에 대한 전체적인 실험 결과를 나타낸다. ‘잡음 추가’ 항목은 원 발성에 잡음을 추가하였을 때의 SNR을 나타낸다. ‘CLEAN’은 잡음을 추가하지 않은 경우를 의미하고, ‘CAR’, ‘SUBWAY’, ‘RESTAURANT’는 각각 Aurora2 DB의 CAR, SUBWAY, RESTAURANT 잡음을 SNR 20dB와 10dB로 삽입한 경우를 의미한다. 이 때, 20dB, 10dB의 기준은 원 발성 대비 추가한 잡음의 크기이므로 실제 SNR는 이보다 더 낮다. ‘baseline’은 특징 강화를 수행하지 않은 경우이며, ‘PCA’, ‘GKPCA’, ‘proposed (VOTE)’, ‘proposed (BORDA)’는 각각 주성분 분석, 그리디 커널 주성분 분석, 제안한 방법 (다수 투표로 결합: VOTE, Borda 계수로 결합: Borda)을 적용한 경우이다.

실험 결과, 제안한 방법은 다른 방법에 비해 평균적으로 더 높은 화자 식별률을 보였다. 추가 잡음이 없는 환경에서는 특징 강화를 수행하지 않았을 때 (‘baseline’) 대체로 가장 높은 화자 식별률을 보였다. 주성분 분석 (‘PCA’)은 추가 잡음이 있을 때 대체로 ‘baseline’보다 식별률이 높았다. 그리디 커널 주성분 분석 (‘GKPCA’)은 추가 잡음이 있는 대부분의 경우에 ‘baseline’보다 높은 식별률을 보였으나 제안한 방법 (‘proposed’)보다는 대체로 저조하였다. ‘baseline’ 실험 대비 상대 오류율 감소량은 주성분 분석 (‘PCA’)이 0.62 %, 그리디 커널 주성분 분석 (‘GKPCA’)이 3.50 %, 다수 투표로 결합한 제안한 방법 (‘proposed (VOTE)’)이 6.97 %, Borda 계수로 결합한 제안한 방법 (‘proposed (BORDA)’)이 6.33 %였다. 그림 7은 제안한 시스템을 구성하는 분류기의 수에 따라 평균 화자

표 2. 실험 결과

Table 2. experimental results.

DB	잡음 추가 (SNR)	잡음 종류	혼합 수	base line	PCA	GK PCA	pro posed (VOTE)	pro posed (BORDA)	최대치	
PC	CLEAN		64	94.10	93.00	92.80	93.50	93.20	94.10	
			256	96.50	96.30	96.30	96.20	96.30	96.50	
	20 dB	CAR	64	64.00	72.90	70.30	74.30	73.70	74.30	
			256	82.90	80.90	85.10	86.10	86.00	86.10	
		SUBWAY	64	50.60	54.10	53.10	53.30	53.00	54.10	
			256	70.30	67.40	67.10	67.80	67.60	70.30	
			64	69.80	70.70	65.90	69.80	68.80	70.70	
			256	80.20	74.80	82.50	83.00	82.70	83.00	
	10 dB	CAR	64	39.30	42.10	40.10	43.30	42.60	43.30	
			256	57.40	40.10	55.00	58.50	57.90	58.50	
		SUBWAY	64	29.00	28.60	29.90	30.10	30.20	30.20	
			256	35.60	33.10	35.70	35.40	35.00	35.70	
64			51.90	48.30	48.00	49.80	49.90	51.90		
256			63.80	61.10	60.70	63.40	62.90	63.80		
휴대 전화	CLEAN		64	75.19	73.08	74.13	75.58	75.48	75.58	
			256	79.90	78.17	78.37	79.62	79.52	79.90	
	20 dB	CAR	64	31.63	35.10	37.69	38.94	38.75	38.94	
			256	45.19	46.63	47.02	46.35	46.63	47.02	
		SUBWAY	64	44.13	47.02	48.65	52.21	51.73	52.21	
			256	57.69	57.21	59.71	60.38	60.00	60.38	
			RESTAURANT	64	46.06	49.42	49.81	52.31	52.40	52.40
				256	53.37	58.56	58.46	59.42	59.04	59.42
	10 dB	CAR	64	20.87	17.40	21.25	22.69	21.92	22.69	
			256	22.21	26.35	28.17	29.33	29.42	29.42	
		SUBWAY	64	28.17	29.90	32.31	35.10	34.04	35.10	
			256	32.21	38.37	38.27	42.12	42.02	42.12	
RESTAURANT			64	27.69	30.77	31.92	34.13	33.46	34.13	
			256	32.02	38.56	39.81	40.96	40.87	40.96	
평균치				52.92	53.21	54.57	56.20	55.90	56.20	

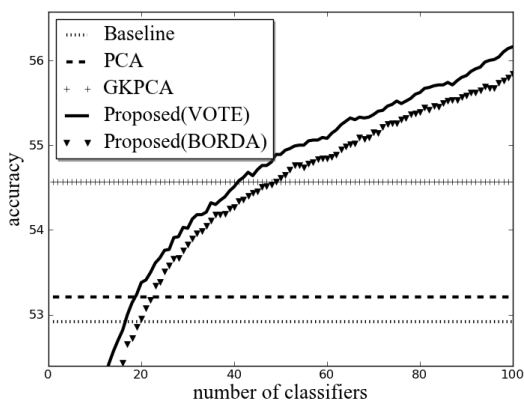


그림 7. 분류기 수에 따른 평균 화자 식별률
Fig. 7. Average speaker identification accuracy according to the number of classifiers.

식별률이 변화하는 것을 나타낸 그래프이다. 이 때, 단일 분류기로 사용했을 경우의 화자 식별률이 높은 순으로 하나씩 수를 늘려 나갔다.

제안한 방법은 그리디 필터링을 사용하는 대신 랜덤하게 부분 특징을 선택하므로 단일 분류기만 사용할 경우 그리디 커널 주성분 분석보다 낮은 식별률을 보인다. 하지만 앙상블 시스템을 구성하는 분류기의 수가 늘어날수록 화자 식별률 역시 증가함을 확인하였다. 제안한 방법(다수투표 방식 기준)은 단일 분류기로 사용 (1개)하였을 때 가장 저조한 39.04%의 화자 식별률을 보였다. 결합한 분류기의 수가 증가할수록 점차 식별률이 증가하다가 17개에서

‘baseline’보다 높아졌고 (53.00%), 41개에서 ‘GKPCA’보다 높아졌으며 (54.58%), 100개를 전부 결합하였을 때 가장 높은 평균 화자 식별률 (56.16%)을 보였다.

V. 실험 설계 및 결과 분석

본 논문에서는 커널 주성분 분석을 이용하여 화자 식별을 수행하는 앙상블 시스템을 제안하였다. 이 시스템은 그리디 커널 주성분 분석과 유사하게 전체 특징으로부터 일부만을 취하여 커널 주성분 분석의 계산량 및 메모리 요구량을 절감하는 정책을 사용하였다. 이 때, 그리디 필터링을 이용하는 대신 반복적으로 일부 특징을 랜덤 선택하고 커널 주성분 분석을 적용하여 복수의 분류기를 학습하였다. 100개의 분류기를 결합한 실험 결과에서 제안한 방법은 다양한 채널 및 잡음 환경에서 평균적으로 가장 높은 화자 식별률을 보였다.

감사의 글

이 논문은 2010년도 정부 (교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행되었습니다. (2010-0024047)

또한 꼼꼼한 검토를 통해 논문상의 많은 오류를 지적해주시고 내용의 질을 개선할 수 있도록 도와주신 한국음향학회지 심사위원분들께 감사드립니다.

참고문헌

1. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2001.
2. B. Scholkopf, A. Smola, K-R. Muller, “Kernel Principal Component Analysis,” in *Proceedings of International Conference on Artificial Neural Networks*, pp. 583-588, 1997.
3. J. Shawe-Taylor, and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
4. 김민석, 양일호, 유하진, “Greedy Kernel PCA를 이용한 화자식별”, *말소리*, 66호, 105-116쪽, 2008.
5. R. Polikar, “Ensemble based Systems in Decision Making,” *Circuits and Systems Magazine*, 2006.
6. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, “Speaker Verification Using Adapted Gaussian Mixture Models,” *Digital Signal Processing*, Vol. 10, pp. 19-41, 2000.
7. D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” *IEEE Transactions on Speech Audio Processing*, Vol. 3, No. 1, pp. 72-83, 1995.
8. H.-G. Hirsch and D. Pearce, “The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions,” in *ASR 2000*, pp. 181-188, 2000.
9. H.-G. Hirsch, *Fant-filtering and noise adding tool*, <http://dnt.kr.hs-niederrhein.de/download.html>.

저자 약력

▶ 양 일 호 (Il-Ho Yang)



2008년: 서울시립대학교 컴퓨터과학부 (학사)
 2010년: 서울시립대학교 컴퓨터과학부 (석사)
 2010년~현재: 서울시립대학교 컴퓨터 과학과 박사과정
 <관심분야> 화자 인식, 음성 인식

▶ 김 민 석 (Min-Seok Kim)



2006년: 서울시립대학교 컴퓨터과학부 (학사)
 2008년: 서울시립대학교 컴퓨터과학부 (석사)
 2011년: 서울시립대학교 컴퓨터과학부 (박사)
 2011년~현재: LG전자 전자기술원 선임 연구원
 <관심분야> 화자 인식, 음성 인식

▶ 소 병 민 (Byung-Min So)



2010년: 서울시립대학교 컴퓨터과학부 (학사)
 2010년~현재: 서울시립대학교 컴퓨터 과학부 석사과정
 <관심분야> 화자인식, 음성인식

▶ 김 명 재 (Myung-Jae Kim)



2010년: 서울시립대학교 컴퓨터과학부 (학사)
 2010년~현재: 서울시립대학교 컴퓨터 과학과 석사과정
 <관심분야> 화자 인식, 음성 인식

▶ 유 하 진 (Ha-Jin Yu)



1990년: KAIST 전산학 (공학사)
 1992년: KAIST 인공지능 (공학석사)
 1997년: KAIST 인공지능 (공학박사)
 1997~2000년: LG전자 전자기술원 선임 연구원
 2000~2002년: SL2 (주) 연구소장
 2002년~현재: 서울시립대학교 컴퓨터 과학부 교수
 <관심분야> 화자 인식, 음성 인식