말소리와 음성과학 제3권 제2호 (2011)
pp. 71~78

특징 강화 방법의 앙상블을 이용한 화자 식별

Speaker Identification Using an Ensemble of Feature Enhancement Methods

양 일 호[1] · 김 민 석[2] · 소 병 민[3] · 김 명 재[4] · 유 하 진[5]
Yang, IL-Ho · Kim, Min-Seok · So, Byung-Min · Kim, Myung-Jae · Yu, Ha-Jin

ABSTRACT

In this paper, we propose an approach which constructs classifier ensembles of various channel compensation and feature enhancement methods. CMN and CMVN are used as channel compensation methods. PCA, kernel PCA, greedy kernel PCA, and kernel multimodal discriminant analysis are used as feature enhancement methods. The proposed ensemble system is constructed with the combination of 15 classifiers which include three channel compensation methods (including 'without compensation') and five feature enhancement methods (including 'without enhancement'). Experimental results show that the proposed ensemble system gives highest average speaker identification rate in various environments (channels, noises, and sessions).

Keywords: classifier ensemble, greedy kernel PCA, kernel multimodal component analysis, speaker identification

1. 서론

화자 식별은 주어진 음성을 누가 발성했는지 파악하는 기술이다. 학습 음성과 테스트 음성의 녹음 채널이 다르거나 잡음이 들어가는 등 환경이 달라지면 화자 식별 시스템의 식별률은 급격히 떨어질 수 있다. 고정된 장소에서만 이용하는 시스템에 대해서는 설치 장소에서 학습 음성을 수집하여 화자 모델을 구성함으로써 채널 불일치 문제가 발생하는 것을 방지할 수 있다. 그러나 전화 통화와 같이 거는 쪽과 받는 쪽의 환경이 고정되지 않은 경우 채널 불일치 문제가 발생하기 쉽다. 녹음한 음성 속에 들어 있는 채널 특성은 받는 쪽의 전화는 물론이고 거는 쪽의 전화에도 영향을 받기 때문이다.

1) 서울시립대학교 heisco@hanmail.net
2) LG전자 전자기술원 minseok3.kim@lge.com
3) 서울시립대학교 sbm1210@naver.com
4) 서울시립대학교 arthmody@naver.com
5) 서울시립대학교 hjyu@uos.ac.kr, 교신저자

이 논문은 2010년도 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2010-0024047)

접수일자: 2011년 5월 30일
수정일자: 2011년 6월 21일
게재결정: 2011년 6월 22일

이러한 사실은 테스트 음성이 어떠한 채널로 녹음되었는지 알기 어려운 음성 과학 수사와 같은 분야에서 화자 식별의 신뢰도를 떨어뜨릴 수 있다. 예를 들어, 전화를 이용하여 이루어지는 범죄의 경우, 범인은 피해자에게 다양한 채널(PC 마이크/일반전화/휴대전화 등)로 전화를 걸어올 수 있다. 이 음성을 화자 식별하기 위해서는 용의자들의 음성을 수집하여 화자 모델을 학습해야 한다. 이 때, 범인의 음성과 용의자의 음성을 녹음한 채널(혹은 그 밖의 환경)이 다르면 화자 식별의 신뢰도를 보장할 수 없다. 실제 상황에서는 범인이 어떠한 채널로 전화를 걸었는지 파악하기 어려울 수 있고, 또한 전송 채널을 파악하더라도 동일한 환경에서 용의자의 음성을 수집하는 것이 현실적으로 어려울 수 있다.

본 연구에서는 학습/테스트 음성의 녹음 채널이 일치하지 않는 상황에서의 화자 식별률 하락 문제를 개선하고자 특징 강화 방법을 이용하였다. 이는 주어진 음성으로부터 추출한 특징을 채널 차이나 잡음 등의 환경에 보다 강인한 특징으로 변환하는 것이다. 그러나 인식 환경을 예측할 수 없는 상태에서 가장 좋은 성능을 보장하는 하나의 채널 보상 방법이나 특징 강화 방법을 선택하기란 매우 어렵다. 서로 다른 특징 강화 방법들은 최적의 성능을 보이는 환경이 각기 다르기 때문이다. 따라서 이러한 특징 강화 방법을 병렬적으로 적용하여 그 결과를 효과적으로 결합하는 방법을 찾고자 하였다.

이를 위해 MFCC(mel-frequency cepstral coefficients) 특징을 다양한 특징 강화 방법으로 변환하고, 각각의 강화된 특징으로 서로 다른 분류기(화자 식별 시스템)를 구성한 뒤 앙상블 결합하였다. 이 때, 특징 강화 방법으로는 주성분 분석(PCA, principal component analysis)[2], 선형 판별 분석(LDA, linear discriminant analysis)[2], 그리디 커널 주성분 분석(GKPCA, greedy KPCA)[1], 커널 다중 판별 분석(KMDA, kernel multimodal component analysis)[4]을 사용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에서 사용한 특징 강화 방법들에 대해 간략히 소개한다. 3장에서는 이러한 특징 강화 방법들로 앙상블을 구성하는 방법을 제안한다. 4장에서 실험 설계 및 결과를 보이고 5장에서 결론을 맺는다.

2. 특징 강화 방법

MFCC는 음성 및 화자 인식 분야에서 가장 널리 쓰이는 특징이다. 본 연구에서는 이를 채널에 강인하게 변환하고자 다양한 특징 강화 방법을 사용하였다. 각각의 특징 강화 방법은 <그림 1>과 같이 수행하였다. 가장 먼저, UBM(universal background model)[7]을 학습하기 위한 음성 특징에 대하여 변환 기저(MFCC 특징을 사상하기 위한 변환 행렬)를 추정한다. 이를 이용해 변환한 특징으로 UBM을 학습하고 MAP[7] 적응을 통해 화자 모델을 구성한다. 테스트를 수행할 때에도 변환된 특징을 이용한다.

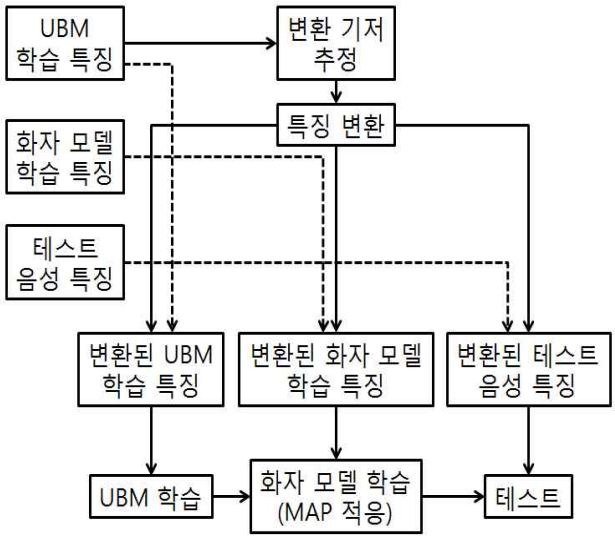

그림 1. 특징 강화 방법.
Figure 1. Feature enhancement method.

2.1 주성분 분석 (PCA)

주성분 분석[2]은 전체 데이터의 분포 정보를 최대한 유지하면서 새로운 기저로 데이터를 사상하는 방법이다.

MFCC 특징에 대해 주성분 분석을 적용하면 화자 식별률이 향상될 수 있다. 그러나 특징이 비선형으로 분포되어 있을 경우, 적합한 변환 기저를 추정하지 못하는 한계를 지닌다.

2.2 선형 판별 분석 (LDA)

선형 판별 분석[2]은 주성분분석과 달리 데이터의 분류 정보를 최대한 유지하면서 새로운 기저로 데이터를 사상하는 방법이다. 본 연구에서는 UBM 학습 데이터를 화자별로 분류하여 선형 판별 분석을 적용하였다.

2.3 그리디 커널 주성분 분석 (GKPCA)

커널 주성분 분석(KPCA)[9]은 커널 방법[10]을 이용하여 비선형 특징을 처리할 수 있도록 개선한 것이다. 커널 방법을 통해 입력 공간(input space)의 원 특징을 고차원의 특징 공간(feature space)으로 사상한 것과 같은 효과를 얻을 수 있다. 이 때, 원 특징이 입력 공간에서 비선형으로 분포되어 있더라도 고차원 특징 공간에서는 선형분리가 가능하게 분포하게 되므로 주성분분석에 비해 좋은 기저를 찾을 수 있다. 그러나, 여기서는 변환 기저를 추정하는데 쓰이는 샘플의 수에 비례하여 계산량 및 메모리 요구량이 크게 증가한다. 음성 및 화자 인식에서는 짧음 발성에서도 많은 수의 특징을 추출하므로, 화자 식별에 커널 주성분 분석을 그대로 적용하기는 어렵다.

그리디 커널 주성분 분석(GKPCA)[1]은 변환 기저를 추정하는데 쓰이는 샘플을 모두 사용하는 것이 아니라, 전체 특징을 대표하는 소수의 부분 집합을 그리디 필터링(greedy filtering)으로 선택하여 커널 주성분 분석 시 발생하는 계산량 및 메모리 요구량을 줄이는 방법이다.

본 연구에서는 식 (1)과 같은 가우시안 RBF 커널 함수를 이용하고 (σ=21), 그리디 필터링으로 선택하는 대표 샘플의 수를 100개로 하였다.

$$k(\vec{x},\vec{y}) = \exp\left(-\frac{\|\vec{x}-\vec{y}\|^2}{2\sigma^2}\right) \quad (1)$$

2.4 커널 다중 판별 분석 (KMDA)

커널 다중 판별 분석[4]은 고차원 특징 공간상에서 서브 클러스터(sub-cluster)의 중심과 전체 중심 간의 거리를 최대화하는 기저로 특징을 사상하는 방법이다. 화자 인식에 적용할 때에는 기저를 추정하기 위한 특징을 각 화자별로 K 개로 군집화한다. <그림 2>는 두 명의 화자(A, B)를 각각 2 개의 서브 클러스터로 군집화한 예이다.

화자가 N명일 경우 총 N×K개의 서브 클러스터를 얻게 된다. 이 방법은 K값이 커질수록 커널 주성분 분석과 유사해진다. 본 연구에서는 UBM 학습 데이터를 화자별로 분류하여 커널 다중

판별 분석을 적용하였다(K=4).

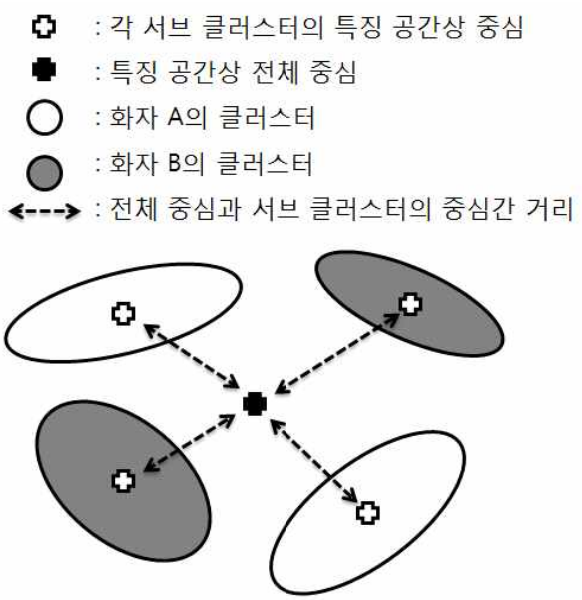

그림 2. KMDA의 목적.
Figure 2. Objective of KMDA

3. 특징 강화 방법의 앙상블

본 연구에서는 2장에서 소개한 특징 강화 방법들을 사용하여 특징을 변환하고 서로 다른 분류기를 학습하여 앙상블[6]을 구성하였다.

3.1 제안한 시스템 구조

<그림 3>는 앙상블 구성의 개략도를 나타내고 있다. 먼저 MFCC로 추출한 UBM 학습 특징으로부터 각각 주성분 분석(PCA), 선형 판별 분석 (LDA), 그리디 커널 주성분 분석(GKPCA), 커널 다중 판별 분석 (KMDA)의 변환 기저를 추정하고 전체 특징을 새로운 기저로 사상한다. 분류기 학습 및 테스트 단계는 각 특징 강화 방법에 대해 2장에서 설명한 <그림 1>을 병렬적으로 수행한다. 변환된 UBM 학습 특징 및 화자 모델 학습 특징으로 UBM을 구성한 후 MAP 적응하여 분류기를 학습한다. 변환된 테스트 특징에 대해 분류 결과를 얻은 후 마지막으로 이 결과들을 앙상블 결합한다.

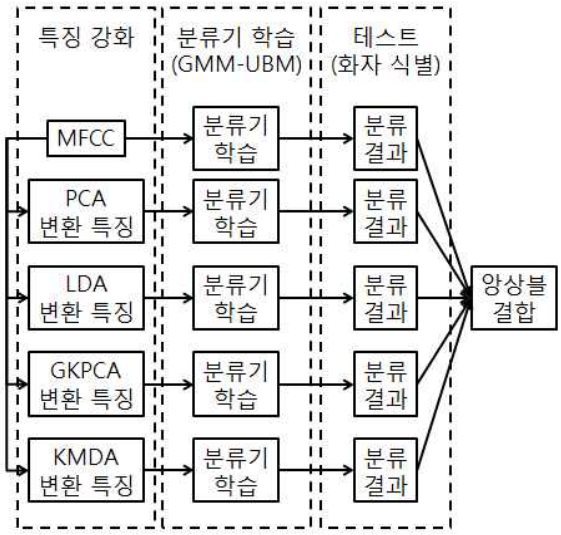

그림 3. 특징 강화 방법의 앙상블.
Figure 3. Ensemble of feature enhancement methods.

3.2 앙상블 결합 방법

3.2.1 다중 투표

다중 투표(majority voting) [6]는 앙상블을 구성하는 한 분류기가 인식한 부류(class)에 대해 1표씩 가산하여 최종적으로 가장 많이 득표한 부류를 인식 결과로 출력한다. T개 분류기를 앙상블 결합하여 C개의 부류(ω) 중 하나를 결과로 취하는 수식은 다음과 같다.

$$J = \arg\max_{j=1,C} \sum_{t=1}^{T} d_{t,j} \qquad (1)$$

이 때, $d_{t,j}$는 t번째 분류기의 결과가 ω_j인지 여부(0 혹은 1)이다. 가장 많은 표를 얻은 부류는 J번째 이므로 앙상블 결합 결과는 ω_J가 된다.

3.2.2 Borda 계수

Borda 계수(Borda count) [6]는 C개 부류에 대한 한 분류기의 인식 결과를 확률이 큰 순서대로 C-1점에서 0점까지 부여하는 투표 방식이다. 본 연구에서는 한 분류기의 인식 결과에서 로그 유사도(likelihood)가 큰 순으로 5개 부류를 취하여 5점에서 1점까지 부여(6번째 부류부터는 0점)하였다. Borda 계수의 수식은 (1)과 동일하나 이 때, $d_{t,j}$는 t번째 분류기의 ω_j에 대한 점수이다.

4. 실험 설계 및 결과

4.1 실험 설계

4.1.1 데이터베이스

다양한 채널 및 잡음 상황에 대해 성능을 평가하기 위해 PC, 일반전화, 휴대전화상에서 녹음한 세 종류의 음성 데이터베이스를 이용하였다. 이는 ETRI 중가마이크 화자인식용 DB, ETRI 화자인식용 일반전화 DB, ETRI 화자인식용 휴대전화 DB 등이다. 각 DB에 대한 실험은 독립적으로 수행하였다. UBM 학습을 위해 월차 화자 전체의 문장 발성 10개(1월차 1회차 발성)를 사용하였고, 화자 모델 학습 및 테스트에는 주차 화자 전체의 문장 발성 10개씩(학습: 1주차 1회차 발성, 같은 시차 테스트: 1주차 3회차 발성, 다른 시차 테스트: 3주차 1회차 발성)을 사용하였다. UBM 학습, 화자 모델 학습, 테스트에 사용한 10개의 문장은 모두 동일한 내용을 발성한 것이다(문장 종속 실험). 이 때 화자 수는 PC DB가 UBM 및 화자 모델 학습에 각각 100명, 일반전화 DB 및 휴대전화 DB가 UBM 학습 101명, 화자 모델 학습 104명이다.

4.1.2 모델 학습

화자 모델 학습에는 GMM-UBM 방법을 사용하였다. 먼저 혼합 수 32개의 GMM으로 UBM을 학습하였다. 이 때, 혼합 수는 1개부터 2배씩 늘려나갔으며 중간 단계의 혼합 수에서는 1회씩 학습하고 혼합 수 32개에서는 총 10회 반복 학습하였다. 각 화자의 학습 발성으로 UBM을 MAP 적응하여 화자 모델을 구성하였다. 이 때, 최대 3회까지 반복 적응을 수행하였다(τ=1).

4.1.3 특징 추출

모든 DB의 sampling rate는 8khz로 일괄 조정하였고, 이로부터 20차 MFCC와 로그 에너지를 특징으로 추출하였다. 특징 레벨에서 에너지를 기반으로 사일런스를 제거하였다. 채널 보상을 위해 발성 별로 각 차원의 평균을 0으로 만드는 CMN(cepstral mean normalization)[3] 및 CMN에 추가하여 분산을 1로 만드는 CMVN(cepstral mean and variance normalization)[11]을 적용하였다. 이를 주성분 분석, 선형 판별 분석, 그리디 커널 주성분 분석, 커널 다중 판별 분석으로 각각 변환하였다.

4.1.4 채널 변환 및 잡음 삽입

학습 데이터와 테스트 데이터의 채널 차이 및 잡음 발생상황에서의 성능을 확인하기 위해 테스트 발성을 전화 채널의 일종인 G.712로 시뮬레이션하고 Aurora2 DB의 자동차 잡음을 SNR 20db로 추가하였다. 채널 시뮬레이션 및 잡음 추가는 FaNT[3]를 이용하였다.

4.2 실험 결과

다음 실험 결과는 화자 모델 학습 시 3회까지 MAP 적응을 수행한 것 중 가장 높은 식별률을 기재한 것이다. ‘MFCC’, ‘CMN’, ‘CMVN’은 각각 MFCC 특징과 이것에 대해 CMN을 적용한 특징, CMVN을 적용한 특징을 의미한다. 특징 강화 방법을 개별적으로 적용한 경우는 ‘[원 특징]-[강화 방법]’으로 표기하였다. 예를 들어, ‘CMN-KMDA’는 CMN을 적용한 MFCC에 대해 커널 다중 판별 분석을 적용한 특징이다. ‘VOTE’와 ‘BORDA’는 특징 강화 방법의 앙상블을 각각 다중 투표 방식과 Borda 계수 방식으로 결합한 것이다. 앙상블을 구성한 경우는 ‘[원 특징]-[결합 방법]’으로 표기하였다. 예를 들어, ‘MFCC-VOTE’는 MFCC 특징을 원 특징으로 하여 개별적으로 특징 강화 방법을 수행한 결과(특징 강화를 적용하지 않은 원 특징을 포함하여 총 5종류)들을 앙상블 결합한 것이다. 마지막으로 ‘TOTAL-[결합 방법]’으로 표기한 것은 앞에서 구한 모든 분류기(3가지 채널 보상 방법과 5가지 특징 강화 방법의 15가지 조합)를 앙상블 결합한 것이다.

4.3 결과 분석

<표 7>은 실험 조건에 따라 평균 식별률을 계산한 것이다. 예를 들어, <표 7>의 ‘동일 채널 평균’은 <표 1>, <표 2>, <표 3>, <표 4>, <표 5>, <표 6>에서 동일 채널 식별률의 평균값을 의미한다. 마찬가지로 모든 실험 중 CLEAN, NOISE, 동일 시차, 상이 시차에 대한 각각의 결과를 평균하여 분석하였다. ‘전체 평균’은 모든 실험 결과의 평균값을 의미한다. 이 중 중요한 몇몇 결과를 <그림 4>의 그래프로 정리하여 비교하였다. <그림 4>는 특징 강화 미적용, 단일 알고리즘, 채널 보상 방법별 앙상블, 전체 앙상블 결과 중 가장 좋은 결과들을 비교한 것이다.

단일 알고리즘만 사용했을 경우에는 평균적으로 CMVN -커널 다중 판별 분석(‘CMVN-KMDA’)이 가장 높은 식별률을 보였다. 채널 보상 방법에 따라 5 종류 특징으로 학습한 분류기를 앙상블 결합하였을 때는 단일 알고리즘만 사용하였을 때보다 평균적으로 더 높은 식별률을 보이는 경우가 있었으나, 경우에 따라 식별률이 달라졌다. 모든 채널 보상 방법과 특징 강화 방법에 대해 15 종류의 분류기를 Borda 계수로 앙상블 결합한 경우(‘TOTAL-BORDA’) 평균적으로 가장 좋은 식별률을 보였다. 이는 상황에 따라 최적의 채널 보상 방법과 특징 강화 방법이 다른데, 각각을 결합하였을 때 상호 보완이 가능하기 때문인 것으로 판단된다.

표 1. 실험 결과 (PC DB, 동일 시차).

Table 1. Experimental results (PC DB, same session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	98.70	71.70	62.60	25.40
CMN	90.40	68.50	85.70	62.00
CMVN	92.80	75.70	89.20	73.80
MFCC-PCA	98.70	72.60	45.50	19.60
CMN-PCA	96.30	72.80	91.50	62.10
CMVN-PCA	95.10	82.20	91.10	75.40
MFCC-GKPCA	**99.00**	74.60	52.40	25.20
CMN-GKPCA	94.80	73.30	90.70	59.40
CMVN-GKPCA	95.60	80.00	92.40	73.40
MFCC-KMDA	98.70	75.10	48.50	23.70
CMN-KMDA	95.30	75.50	91.90	67.00
CMVN-KMDA	95.50	82.50	92.80	79.50
MFCC-LDA	98.40	72.10	36.50	19.10
CMN-LDA	95.10	78.10	90.80	62.40
CMVN-LDA	95.10	78.80	92.10	73.10
MFCC-VOTE	**99.00**	76.70	52.40	25.80
MFCC-BORDA	95.20	77.80	92.30	66.70
CMN-VOTE	96.10	83.50	93.90	**80.60**
CMN-BORDA	98.80	76.20	50.50	24.10
CMVN-VOTE	94.60	76.40	92.00	65.20
CMVN-BORDA	96.10	83.10	94.00	79.00
TOTAL-VOTE	97.60	87.60	**95.20**	79.30
TOTAL-BORDA	97.20	**89.10**	93.40	80.10

표 2. 실험 결과 (PC DB, 상이 시차).

Table 2. Experimental results (PC DB, difference session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	86.20	50.90	41.10	17.50
CMN	76.50	51.20	68.00	48.10
CMVN	80.30	61.70	72.80	56.20
MFCC-PCA	87.70	53.70	28.10	13.40
CMN-PCA	85.50	58.80	74.90	46.40
CMVN-PCA	83.00	66.50	74.80	57.40
MFCC-GKPCA	85.70	55.10	32.90	16.80
CMN-GKPCA	83.10	57.20	74.70	46.00
CMVN-GKPCA	82.50	66.90	74.30	58.40
MFCC-KMDA	86.80	54.10	29.80	14.10
CMN-KMDA	83.80	58.40	75.40	45.50
CMVN-KMDA	83.60	67.70	76.00	60.40
MFCC-LDA	86.40	53.90	23.60	12.60
CMN-LDA	82.60	61.20	73.30	47.40
CMVN-LDA	84.20	65.20	79.10	54.80
MFCC-VOTE	**88.30**	56.00	33.20	15.90
MFCC-BORDA	84.50	61.00	76.90	50.10
CMN-VOTE	84.90	68.80	80.60	**61.70**
CMN-BORDA	87.80	54.90	31.80	16.40
CMVN-VOTE	84.00	60.30	76.60	50.90
CMVN-BORDA	84.30	68.20	79.50	60.80
TOTAL-VOTE	88.20	**72.30**	**81.70**	59.50
TOTAL-BORDA	87.30	71.70	79.40	61.40

표 3. 실험 결과 (일반전화 DB, 동일 시차).

Table 3. Experimental results (phone DB, same session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	87.02	36.06	69.33	37.88
CMN	82.31	30.67	79.71	33.56
CMVN	83.27	46.44	81.15	47.88
MFCC-PCA	87.50	36.06	56.54	15.87
CMN-PCA	88.17	29.71	85.96	33.56
CMVN-PCA	88.08	51.63	84.33	52.98
MFCC-GKPCA	87.12	37.69	65.00	32.69
CMN-GKPCA	89.23	34.33	87.12	40.10
CMVN-GKPCA	88.56	52.98	85.29	54.71
MFCC-KMDA	87.40	36.92	71.73	36.06
CMN-KMDA	89.52	27.69	86.92	33.75
CMVN-KMDA	89.13	54.33	86.73	56.06
MFCC-LDA	87.21	27.98	52.98	26.35
CMN-LDA	89.13	35.29	87.12	40.10
CMVN-LDA	88.46	54.71	85.48	55.10
MFCC-VOTE	87.98	40.19	65.38	33.27
MFCC-BORDA	89.33	32.50	87.50	37.79
CMN-VOTE	89.23	54.81	86.63	55.77
CMN-BORDA	87.69	39.33	62.60	28.94
CMVN-VOTE	88.85	32.12	86.73	38.94
CMVN-BORDA	88.85	54.71	86.35	55.67
TOTAL-VOTE	**93.08**	56.63	**89.42**	56.83
TOTAL-BORDA	92.40	**61.63**	87.60	**61.73**

표 4. 실험 결과 (일반전화 DB, 상이 시차).

Table 4. Experimental results (phone DB, difference session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	65.00	26.83	44.04	27.79
CMN	56.44	17.69	50.10	21.54
CMVN	58.56	29.42	53.37	30.19
MFCC-PCA	68.85	24.13	35.19	22.79
CMN-PCA	66.06	20.58	61.44	22.60
CMVN-PCA	63.94	32.40	58.65	32.60
MFCC-GKPCA	66.54	25.38	40.58	23.65
CMN-GKPCA	66.06	24.81	62.12	27.12
CMVN-GKPCA	63.75	32.79	58.56	32.69
MFCC-KMDA	66.92	20.96	44.52	25.87
CMN-KMDA	66.54	19.81	62.02	24.13
CMVN-KMDA	64.71	32.69	59.42	33.46
MFCC-LDA	67.02	20.87	32.60	19.71
CMN-LDA	62.98	23.17	59.13	26.63
CMVN-LDA	63.85	34.23	58.56	34.33
MFCC-VOTE	67.40	28.37	40.48	25.19
MFCC-BORDA	66.44	23.08	62.12	26.83
CMN-VOTE	65.29	34.13	60.67	33.56
CMN-BORDA	67.50	27.88	39.13	24.23
CMVN-VOTE	64.62	24.04	60.67	27.12
CMVN-BORDA	64.71	34.52	59.71	33.56
TOTAL-VOTE	**70.10**	37.40	**63.46**	36.73
TOTAL-BORDA	69.90	**41.54**	62.31	**42.50**

표 5. 실험 결과 (휴대전화 DB, 동일 시차).

Table 5. Experimental results (cellphone DB, same session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	88.27	18.94	64.04	29.62
CMN	81.15	35.58	78.17	37.98
CMVN	80.38	52.31	76.73	56.35
MFCC-PCA	88.75	5.19	50.67	24.81
CMN-PCA	86.35	10.96	83.46	35.00
CMVN-PCA	82.88	48.75	78.08	57.12
MFCC-GKPCA	88.27	9.62	54.04	23.85
CMN-GKPCA	85.48	26.73	83.46	38.46
CMVN-GKPCA	82.31	48.17	78.94	57.98
MFCC-KMDA	88.27	9.81	63.94	29.04
CMN-KMDA	86.63	35.00	83.94	46.35
CMVN-KMDA	84.13	48.85	79.23	58.17
MFCC-LDA	88.46	3.85	46.92	22.21
CMN-LDA	85.10	25.67	81.44	36.35
CMVN-LDA	84.13	42.79	80.10	54.90
MFCC-VOTE	**90.38**	8.56	55.87	27.50
MFCC-BORDA	87.60	31.63	84.71	42.02
CMN-VOTE	85.38	52.50	80.77	60.38
CMN-BORDA	90.19	8.46	53.85	26.06
CMVN-VOTE	86.15	29.52	84.42	41.83
CMVN-BORDA	85.19	**53.37**	80.00	**61.44**
TOTAL-VOTE	90.00	40.67	**85.19**	58.56
TOTAL-BORDA	88.65	46.06	82.60	57.88

표 6. 실험 결과 (휴대전화 DB, 상이 시차).

Table 6. Experimental results (cellphone DB, difference session).

	동일 채널		상이 채널	
특징	CLEAN	NOISE	CLEAN	NOISE
MFCC	61.06	9.33	38.65	21.73
CMN	56.73	24.23	54.13	25.29
CMVN	57.79	**36.92**	53.94	36.63
MFCC-PCA	63.94	2.31	29.90	14.90
CMN-PCA	63.46	18.75	59.04	22.88
CMVN-PCA	60.77	32.02	55.10	38.17
MFCC-GKPCA	62.40	6.83	30.38	15.19
CMN-GKPCA	61.35	18.17	57.88	25.10
CMVN-GKPCA	60.87	31.06	56.63	38.85
MFCC-KMDA	64.71	7.40	42.69	19.90
CMN-KMDA	63.85	23.37	60.19	30.77
CMVN-KMDA	60.67	31.25	56.54	38.17
MFCC-LDA	63.75	2.69	27.88	15.00
CMN-LDA	61.83	18.17	58.75	24.81
CMVN-LDA	62.88	28.17	56.92	37.02
MFCC-VOTE	65.29	5.29	33.85	18.37
MFCC-BORDA	63.94	22.21	60.58	27.88
CMN-VOTE	63.46	34.23	58.56	41.06
CMN-BORDA	64.81	5.48	32.79	17.21
CMVN-VOTE	63.56	20.96	60.87	27.31
CMVN-BORDA	63.56	34.52	57.98	40.77
TOTAL-VOTE	**69.13**	26.44	**61.06**	39.52
TOTAL-BORDA	68.17	32.50	59.71	**41.25**

표 7. 결과 분석.

Table 7. Analysis results.

특징	동일 채널 평균	상이 채널 평균	CLEAN 평균	NOISE 평균	동일 시차 평균	상이 시차 평균	전체 평균
MFCC	58.33	39.97	67.17	31.14	57.46	40.84	49.15
CMN	55.95	53.69	71.61	38.03	63.81	45.83	54.82
CMVN	62.97	60.69	73.36	50.30	71.33	52.32	61.83
MFCC-PCA	57.45	29.77	61.78	25.45	50.15	37.08	43.61
CMN-PCA	58.12	56.57	78.51	36.18	64.66	50.03	57.35
CMVN-PCA	65.61	62.98	76.32	52.26	73.97	54.61	64.29
MFCC-GKPCA	58.19	34.39	63.69	28.88	54.12	38.45	46.29
CMN-GKPCA	59.55	57.68	78.00	39.23	66.93	50.30	58.61
CMVN-GKPCA	65.46	63.51	76.64	52.33	74.20	54.78	64.49
MFCC-KMDA	58.09	37.49	66.17	29.41	55.76	39.81	47.79
CMN-KMDA	60.45	58.99	78.83	40.61	68.29	51.15	59.72
CMVN-KMDA	66.26	64.71	77.37	53.59	75.58	55.38	65.48
MFCC-LDA	56.05	27.95	59.31	24.70	48.51	35.50	42.00
CMN-LDA	59.86	57.35	77.27	39.94	67.22	50.00	58.61
CMVN-LDA	65.21	63.46	77.57	51.10	73.73	54.94	64.33
MFCC-VOTE	59.46	35.60	64.96	30.10	55.25	39.80	47.53
MFCC-BORDA	61.27	59.62	79.26	41.63	68.76	52.13	60.44
CMN-VOTE	67.69	66.18	78.79	55.09	76.63	57.25	66.94
CMN-BORDA	59.09	33.97	63.96	29.10	53.89	39.16	46.53
CMVN-VOTE	60.43	59.38	78.59	41.22	68.06	51.75	59.90
CMVN-BORDA	67.59	65.73	78.35	54.97	76.48	56.84	66.66
TOTAL-VOTE	69.10	67.21	**82.01**	54.29	77.51	58.80	68.15
TOTAL-BORDA	**70.51**	**67.49**	80.72	**57.28**	**78.20**	**59.81**	**69.00**

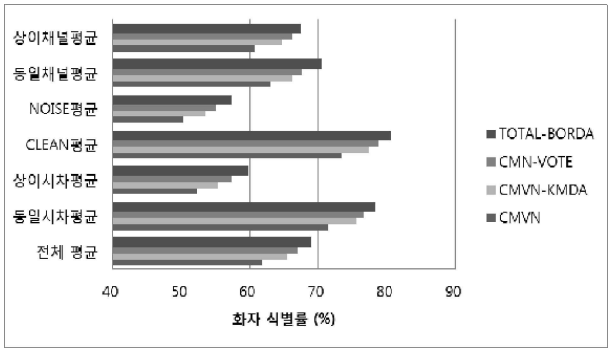

그림 4. 주요 알고리즘들의 화자 식별률.

Figure 4. Speaker identification rate of major algorithms.

5. 결론

음성 과학 수사와 같이 특수한 분야에서는 학습 음성과 테스트 음성의 녹음 환경(채널, 잡음, 시차 등)이 다르고, 그러한 환경 차이에 대한 사전 정보를 얻기 어려울 수 있다. 이러한 상황에서는 화자 식별 시스템의 식별률이 저하되므로 식별 결과의 신뢰도를 보장할 수 없다. 본 연구에서는 이러한 문제를 특징 강화 방법으로 개선하고자 하였다. 하지만 모든 상황에서 최적의 성능을 보이는 하나의 특징 강화 방법은 존재하지 않으므로, 여러 방법을 결합하여 평균적인 식별률을 개선하고자 하였다.

이를 위해 MFCC 특징에 CMN, CMVN과 같은 채널 보상 방법과 주성분 분석, 선형 판별 분석, 그리디 커널 주성분 분석, 커널 다중 판별 분석과 같은 특징 강화 방법을 각각 적용하여 서로 다른 분류기를 학습하고 앙상블 결합하는 방법을 제안하였다. 다양한 상황에서의 화자 식별률을 확인하기 위해 PC, 일반전화, 휴대전화 상에서 수집한 음성에 대하여 채널, 잡음, 시차 등을 달리하여 실험하였다. 실험 결과 제안한 앙상블 방법은 평균적으로 가장 높은 식별률을 보였다.

참고문헌

[1] Kim, M-S., Yang, I-H., Yu, H-J. (2008). “Speaker Identification using Greedy Kernel PCA”, *Malsori*, No. 66, 105-116.
(김민석, 양일호, 유하진, (2008). “Greedy Kernel PCA를 이용한 화자식별”, 말소리, 66호, 105-116.)

[2] Duda, R. O., Hart, P. E., Stork, D. G. (2001). *Pattern Classification*, New York: John Wiley & Sons.

[3] Hirsch, H. -G. (2011). “FaNT - Filtering and Noise Adding Tool”, http://dnt.kr.hs-niederrhein.de/download.html.

[4] Kim, M-S., Yang, I-H., Yu, H-J. (2010). “Kernel Multimodal Discriminant Analysis for Speaker Verification”, In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4498-4501.

[5] Liu, F., Stern, R., Huang, X., Acero, A. (1993) “Efficient Cepstral Normalization for Robust Speech Recognition”, In *Proceedings of ARPA Speech Natural Language Workshop*, 69-74.

[6] Polikar, R. (2006). “Ensemble based Systems in Decision Making”, *Circuits and Systems Magazine*, Vol. 6, Issue 3, 21-45.

[7] Reynolds, D. A., Quatieri, T. F., Dunn, R. B. (2000). “Speaker Verification Using Adapted Gaussian Mixture Models”, *Digital Signal Processing*, Vol. 10, 19-41.

[8] Reynolds, D. A., Rose, R. C. (1995). “Robust text-independent speaker identification using Gaussian mixture speaker models”, *IEEE Transactions on Speech Audio Processing*, Vol. 3, No. 1, 72-83.

[9] Scholkopf, B., Smola, A., Muller, K-R. (1997). “Kernel Principal Component Analysis”, In *Proceedings of International Conference on Artificial Neural Networks*, 583-588.

[10] Shawe-Taylor, J., Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge: Cambridge University Press.

[11] Viikki, O., Laurila, K. (1998). “Cepstral Domain Segmental Feature Vector Normalization for Noise Robust Speech Recognition”, *Speech Communication*, Vol. 25, Issues 1-3, 133-147.

• **양일호 (Yang, IL-Ho)**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: hesico@hanmail.net
관심분야: 음성인식, 화자인식
현재 컴퓨터과학부 대학원 박사과정 재학중

• **김민석 (Kim, Min-Seok)**
LG전자 전자기술원
서울시 서초구 양재동 221
Email: minseok3.kim@lge.com
관심분야: 음성인식, 화자인식

• **소병민 (So, Byung-Min)**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: sbm1210@naver.com
관심분야: 음성인식, 화자인식
현재 컴퓨터과학부 대학원 석사과정 재학중

• **김명재 (Kim, Myung-Jae)**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: arthmody@naver.com
관심분야: 음성인식, 화자인식
현재 컴퓨터과학부 대학원 석사과정 재학중

• **유하진 (Yu, Ha-Jin)**, **교신저자**
서울시립대학교 컴퓨터과학부
서울시 동대문구 전농동 90번지
Tel: 02-2210-5322 Fax: 02-2210-5275
Email: hjyu@uos.ac.kr
관심분야: 음성인식, 화자인식
2002~현재 컴퓨터과학부 교수