

Glottal flow 신호에서의 향상된 특징추출 및 다중 특징파라미터 결합을 통한 화자인식 성능 향상

강지훈 · 김영일 · 정상배*

Performance Improvement of Speaker Recognition Using Enhanced Feature Extraction in Glottal Flow Signals and Multiple Feature Parameter Combination

Jihoon Kang · Youngil Kim · Sangbae Jeong*

Department of Electronics Engineering/ERI, Gyeongsang National University, Jinju 660-701, Korea

요 약

본 논문에서는 화자 인식의 성능을 개선하기 위해서 glottal flow로부터 source mel-frequency cepstral coefficient (SMFCC), 왜도, 첨도를 추출하여 활용하였다. 일반적으로 glottal flow의 고주파 대역은 응답의 크기가 평탄하므로 미리 정한 차단주파수 미만에 대해서만 SMFCC를 추출한다. 추출된 SMFCC, 왜도, 첨도는 종래의 특징 파라미터와 결합된 후 종래의 화자인식 시스템과 동등한 조건에서의 성능 비교를 위하여 principal component analysis (PCA) 및 linear discriminant analysis (LDA)를 통한 차원축소가 행해진다. 대용량의 화자인식 실험결과를 통해서 제안된 인식 시스템이 종래의 화자인식 시스템 보다 더 좋은 성능을 나타냄을 확인할 수 있었으며, 특히 가우시안 혼합이 낮을 때 더 높은 성능향상을 나타내었다.

ABSTRACT

In this paper, we utilize source mel-frequency cepstral coefficients (SMFCCs), skewness, and kurtosis extracted in glottal flow signals to improve speaker recognition performance. Generally, because the high band magnitude response of glottal flow signals is somewhat flat, the SMFCCs are extracted using the response below the predefined cutoff frequency. The extracted SMFCC, skewness, and kurtosis are concatenated with conventional feature parameters. Then, dimensional reduction by the principal component analysis (PCA) and the linear discriminant analysis (LDA) is followed to compare performances with conventional systems under equivalent conditions. The proposed recognition system outperformed the conventional system for large scale speaker recognition experiments. Especially, the performance improvement was more noticeable for small Gaussian mixtures.

키워드 : 화자인식, 성분파, 왜도, 첨도, 주성분 분석, 주요인 분석

Key word : speaker recognition, glottal flow, skewness, kurtosis, PCA, LDA

Received 22 October 2015, Revised 08 November 2015, Accepted 23 November 2015

* Corresponding Author Sangbae Jeong(E-mail:jeongsb@gnu.ac.kr, Tel:+82-55-772-1727)

Department of Electronics Engineering/ERI, Gyeongsang National University, Jinju 660-701, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2015.19.12.2792>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

최근 정보 통신망을 이용한 전자 정부, 전자 상거래, 개인 맞춤 서비스 등의 수요가 급격히 증가하면서, 본인 인증절차가 필수가 되었다. 하지만, 기존의 본인인증절차에 필요한 스마트카드, PIN 등은 분실의 위험이 있으며, 주기적으로 변경해주어야 한다는 단점을 갖고 있다. 이러한 단점을 보완하기 위하여 생체인식기술이 각광받고 있다. 특히, 생체인식기술중 화자인식은 지문인식, 홍채인식에 비해 비용경쟁력이 높고 원격인증이 가능하다는 장점으로 인해 지속적으로 연구가 수행되고 있다[1].

화자인식이란 사람이 발생하는 음성신호에서 고유한 특색을 추출하여 화자를 확인하는 것으로 정의되며, 화자인식 연구는 크게 특징추출과 패턴 매칭으로 구성되어 있다. 본 논문에서는 특징추출과 관련된 연구를 중점적으로 수행하였다. 현재, 화자인식을 위해서 가장 널리 사용되는 특징은 mel-frequency cepstral coefficient (MFCC), perceptual linear predictive cepstral coefficient (PLPCC) 등이 있다[2,12]. 하지만, 이와 같은 특징들은 음성의 형태학적 특성을 고려하지 않고 있으며 화자가 발생하는 어휘에 영향을 많이 받는다는 단점을 갖고 있다[4,6]. 이를 보완하기 위하여 본 논문에서는 MFCC, PLPCC와 함께 음성의 형태학적 특성을 포함하고 있는 glottal flow에서 SMFCC, 왜도, 첨도 특징을 추출하여 화자인식에 사용하였다. 또한, glottal flow 스펙트럼의 고주파대역에서 발생하는 평탄한 성분은 화자정보를 추출하는데 무의미하므로 저역통과필터링 및 스펙트럼 확장을 수행한 후 특징을 추출하였다. 최종적으로 추출된 특징들은 결합 후 PCA 와 LDA 과정을 통해 차원축소를 수행하여 화자인식에 사용하였다.

본 논문의 구성은 다음과 같다. 2장에서 화자인식 관련 연구를 설명하고, 3장에서 제안된 화자인식 시스템에 대하여 소개한다. 4장에서 실험방법 및 실험 결과를 분석하고 마지막으로 5장에서 결론을 맺는다.

II. 관련 연구

2.1. MFCC 특징 파라미터 추출

MFCC는 화자인식에 사용되는 대표적인 파라미터로서 사람의 청각특성을 반영한 비 균일한 스케일로

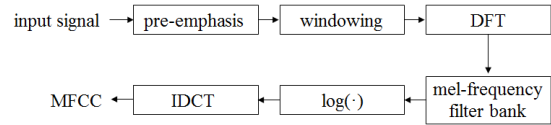


Fig. 1 Process of MFCC extraction

주파수를 나눌 수 있다는 장점을 갖고 있다[2]. MFCC를 추출하는 과정을 그림 1에 나타내었다.

입력 신호는 입에서 자유공간으로 방사하는 과정에서 고주파 성분이 손실된다. 손실된 고주파 성분을 보상해주기 위하여 식 (1)로 정의된 선강조(pre-emphasis) 필터를 수행한다. 선강조 필터를 통과한 입력신호는 원활한 신호 분석을 위하여 창함수(window)를 씌워서 신호를 프레임 단위로 나누어 분석한다.

$$H(z) = 1 - 0.97z^{-1} \quad (1)$$

그 후, discrete Fourier transform (DFT)를 이용하여 주파수 영역의 응답으로 변환되며 응답의 절대치를 인간의 청각 특성을 반영한 멜-주파수 필터뱅크로 적분한다. 선형주파수와 멜-주파수의 관계는 식 (2)와 같다.

$$Mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (2)$$

최종적으로 로그를 취한 필터뱅크 응답에 inverse discrete cosine transform (IDCT)를 취하여 최종적으로 MFCC 특징 파라미터를 추출한다[3].

2.2. Gaussian Mixture Model (GMM)

GMM은 화자인식에 가장 많이 사용되는 확률분포 모델링 기법으로서 식 (3)과 같이 표현된다[4].

$$P(\mathbf{x}|\lambda) = \sum_{i=1}^M w_i b_i(\mathbf{x}) \quad (3)$$

\mathbf{x} 는 D차 특징 벡터, w_i 는 i 번째 가우시안 혼합의 가중치로서 가중치의 총 합은 항상 1을 만족한다. M은 가우시안 혼합의 개수, $b_i(\mathbf{x})$ 는 D차원의 가우시안 분포로서 식 (4)로 정의된다.

$$b_i(\mathbf{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} \quad (4)$$

μ_i 는 i 번째 가우시안 분포의 평균벡터, Σ_i 는 i 번째 가우시안 분포의 공분산행렬이다. λ 는 가우시안 분포의 평균벡터, 공분산행렬, 가중치 등의 파라미터의 집합을 나타낸다.

가우시안 혼합 모델의 파라미터는 주어진 훈련 데이터의 우도의 총합이 최대가 되도록 하는 방식, 즉, expectation maximization (EM) 기법으로 추정된다. EM 알고리즘은 초기 파라미터 λ 를 설정하고 $P(x|\bar{\lambda}) \geq P(x|\lambda)$ 를 만족하는 새로운 파라미터 $\bar{\lambda}$ 를 추정한다. 가우시안 혼합 모델의 파라미터 초기값은 일반적으로 벡터 양자화 기법을 이용하여 추정하며, 새로운 파라미터는 다음 반복의 초기 파라미터가 현재파라미터와 초기파라미터의 차가 특정 수렴 값에 도달하거나 최대 반복 횟수를 만족 할 때까지 반복하여 추정된다[5].

2.3. Glottal flow 추정

glottal flow란 성대의 진동에 의해 분절적으로 만들어지는 작은 공기 파열을 의미하며 화자의 형태학적인 특성을 반영하고 있다. glottal flow는 일반적으로 Iterative adaptive inverse filtering (IAIF)을 통해 추정된다[6]. IAIF는 정확한 glottal flow추정을 위하여 2 단계의 과정을 거친다. 상세한 과정을 그림 2에 나타내었다. 그림 2에서 $a_r(k)$ 는 r 차수를 갖는 linear predictive coefficient (LPC)의 k 번째 값이다. $g(n)$ 은 최종적으로 추정되는 glottal flow이다. 각 블록에 대한 설명은 다음과 같다. 고역통과 필터링은 음성을 녹음하는 과정에서 발생하는 저주파 전원 잡음을 제거하기 위하여 수행된다. 적분은 입술에서 자유공간으로 방사하는 과정에서 발생하는 방사효과를 제거하며 전달함수는 식 (5)와 같다.

$$H(z) = \frac{1}{1 - 0.97z^{-1}} \quad (5)$$

LPC 분석은 성도 특성을 모델링하기 위하여 수행되며, 입력신호의 자기상관함수에 Levinson-Durbin (LD) 재귀 알고리즘을 수행하여 추출한다[7]. 마지막으로 역 필터링은 LPC를 사용하여 음성의 성도 특성을 제거하며 이를 위한 전달함수는 식 (6)과 같다.

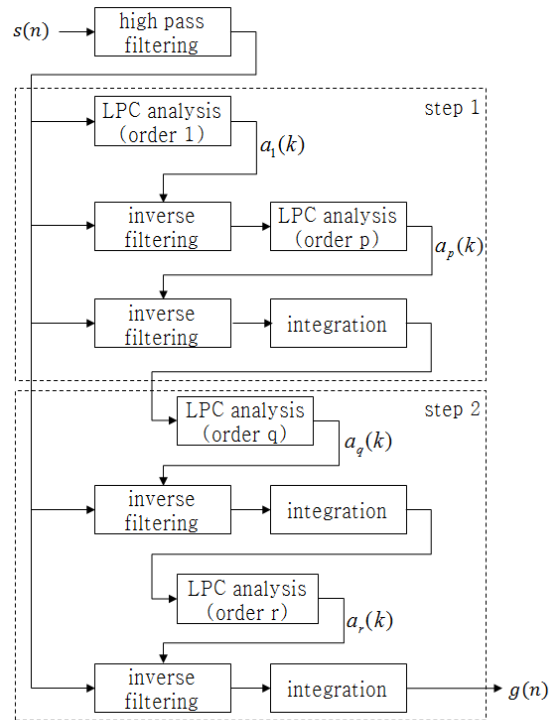


Fig. 2 Glottal flow estimation using the IAIF

$$Q(z) = \sum_{k=1}^r a_r(k) z^{-k} \quad (6)$$

최종적으로 추정된 glottal flow에는 입력음성의 스펙트럼에서 성도의 특성이 제거된 오로지 화자의 형태학적인 특성으로 인한 성분만을 포함하게 된다[8].

2.4. 왜도 및 첨도

음성은 성대의 떨림 여부에 따라 유성음과 무성음으로 구분되며 무성음의 경우 성대가 열린 상태에서 진동 없이 소리가 발생된다. 따라서 무성음은 시간 영역에서 신호의 분포가 정규분포에 가깝다고 예상할 수 있으며 이는 화자인식을 위한 특징파라미터로 사용하기에는 적합하지 않다[9]. 반면, 유성음의 경우 성대의 떨림으로 인해 주기적인 펄스 형태를 갖는 소리가 발생한다. 즉, 유성음은 시간 영역에서의 분포가 정규분포보다 뾰족하거나 왼쪽 혹은 오른쪽으로 상대적으로 치우쳐 있을 것으로 예상할 수 있다. 이러한 유성음의 분포 특성은 성대의 떨림 특성과 관련이 있으며 화자마다 같지 않을 것이

라고 예상할 수 있으므로 특징파라미터로서 활용될 수 있다. 유성음 표본에 관한 분포 특성을 분석하는 방법으로 왜도와 첨도가 있다. 왜도는 평균을 기준으로 표본 집합의 분포가 대칭일 경우에 0, 오른쪽으로 치우칠 때 양수, 왼쪽으로 치우칠 때 음수 값을 갖는다. 첨도는 표본 집합의 분포의 뾰족함을 나타낸다. 정규분포일 경우에 0의 값을 가지며 분포가 상대적으로 뾰족해질수록 더 큰 양의 값을 가진다. 왜도 및 첨도는 식 (7)과 같이 각각 3차 혹은 4차 모멘트로부터 추정 가능하[10].

$$S = \frac{\sum_{n=0}^{N-1} (x(n) - \mu)^3}{(N-1)\sigma^3}, K = \frac{\sum_{n=0}^{N-1} (x(n) - \mu)^4}{(N-1)\sigma^4} - 3 \quad (7)$$

N 은 표본의 총 개수, μ 는 주어진 표본 집합의 평균, $x(n)$ 은 n 번째 표본의 값, σ 는 표본 집합이 갖는 표준편차이다.

2.5. PCA 및 LDA

PCA와 LDA는 데이터 분류와 차원 축소를 위하여 가장 널리 사용되는 기술이다. PCA는 주어진 데이터간의 상관관계로 인한 정보손실을 최소화하여 데이터의 분산을 최대화 하는 주성분을 찾아 각 데이터를 주성분 축으로 사영시키는 방법으로서 데이터의 분포를 가장 잘 나타내는 다변량 통계 기법이다. 반면, LDA는 데이터의 클래스 간 분산과 클래스 내 분산의 비율을 최대화하여 데이터의 클래스 분리를 최대화하는 축을 찾아

데이터를 사영시키는 방법이다[11]. PCA, LDA는 식 (8), (9)를 수행하여 고유값과 고유벡터를 구하고 고유값이 큰 고유벡터를 축으로 하여 데이터를 사영시켜 차원을 축소한다.

$$Re = \lambda e \quad (8)$$

$$S_B e = \lambda S_W e \quad (9)$$

R 은 입력된 데이터의 공분산 행렬, S_B 는 클래스 간 분산 행렬, S_W 는 클래스 내 분산 행렬, λ 는 고유값, e 는 고유벡터이다.

III. 제안된 화자 인식 시스템

본 논문에서 제안한 화자인식 시스템의 블록다이어그램을 그림 3에 나타내었다. 단구간 음성이 입력되었을 때, 불필요한 정보를 제거하기 위하여 단순 에너지 임계치를 적용하여 묵음제거를 수행한다. 그 후, PLPCC, MFCC를 추출하고 glottal flow 신호로부터 제안된 방식에 의한 SMFCC, 왜도, 첨도 등을 추가적으로 추출한다. 추출된 특징 파라미터들은 결합 후, PCA 및 LDA에 의한 차원축소를 수행하여 화자인식에 사용한다. 최종 추출된 특징 파라미터는 각각의 화자에 관한 미리 훈련된 GMM에 입력하여 특정 화자에 대한 유사도가 측정된다. 주요 블록에 대한 설명은 다음과 같다.

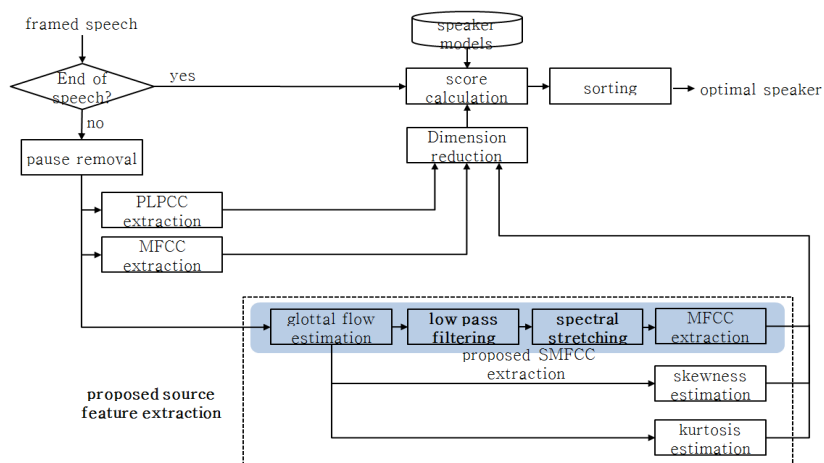


Fig. 3 Proposed speaker recognition system

3.1. 묵음제거

묵음제거는 입력음성에서 화자정보를 포함하고 있지 않은 데이터를 제거함으로써 계산량 감축 및 화자인식 신뢰성을 높이기 위하여 수행된다. 본 논문에서는 초기 입력음성의 평균 에너지를 사용하여 임계값을 구하고 단구간 프레임에서의 평균에너지가 임계값보다 클 경우 음성작을 경우 묵음으로 분류하여 묵음제거를 수행하였다.

3.2. PLPCC 추출

음성의 주파수 영역의 특성 분석을 위하여 MFCC의 보조적인 특징 파라미터로서 PLPCC가 널리 활용되어 왔다[12]. PLPCC는 인지적 필터뱅크를 활용한 주파수 응답의 추정, 인지적 관점에서의 주파수 응답 크기 변환, 주파수 응답으로부터의 LPC의 추정, LPC로부터의 캡스 트럼 계수의 추정 등의 과정을 통해서 추출된다[13].

3.3. 제안된 음원 특징 추출

일반적으로 glottal flow의 스펙트럼은 그림 4의 (a)와 같이 고주파 대역에서 응답의 크기가 평탄해지는 것을 확인할 수 있다. 이러한 평탄한 스펙트럼 부분에서의 정보 추출은 무의미하기 때문에 저역통과 필터링을 수행하여 제거하여 준다. 차단 주파수의 설정 문제는 고대역에서 glottal flow 스펙트럼이 갖는 평탄함 정도가 유사할 것이라고 단정할 근거가 없으므로 전수조사를 통하여 결정한다. 또한, 종래에 구현해 놓은 MFCC 추출 블록을 동일하게 활용하여 SMFCC를 추출하기 위하여 스펙트럼 확장을 수행한다. 확장 과정에서 발생하는 정의되지 않은 스펙트럼의 값은 이미 정의 되어 있는 스펙트럼 값의 선형 보간에 의해서 간단히 추정한다. 그림 4는 차단 주파수가 6 kHz 일 때의 선형 보간 추정법을 활용한 glottal flow 스펙트럼의 확장 예시를 나타내었다.

3.4. 차원 감축

본 논문에서 활용하는 최종적인 특징 파라미터는 기존 화자인식에서 널리 사용되어온 MFCC 및 PLPCC, 그리고 glottal flow에서 추출한 SMFCC, 그것의 왜도 및 첨도로 구성된다. 상기의 특징 파라미터는 식 (10)의 변환을 통하여 차원 축소가 이루어지며 변환된 벡터가 화자인식을 위한 특징 파라미터로 활용된다.

$$c = Ax \quad (10)$$

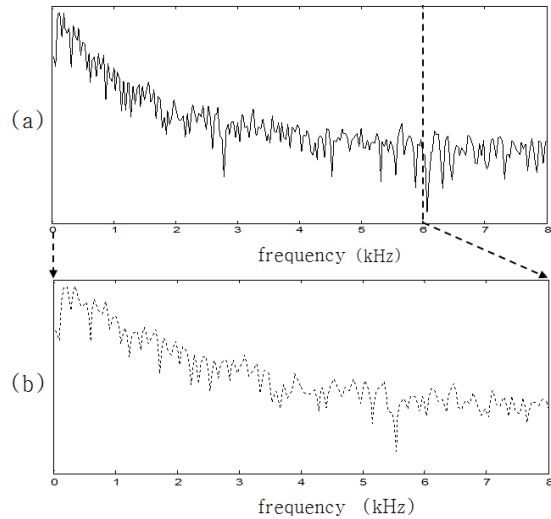


Fig. 4 Example of original and stretched spectrum of glottal flow ((a) original (b) stretched)

이때, $x = [c_{MFCC}; c_{PLPCC}; c_{SMFCC}; S; K]$, A 는 PCA 혹은 LDA에 의해 추정된 차원축소 행렬이다. c_{MFCC} , c_{PLPCC} , c_{SMFCC} , S , K 는 각각 MFCC 특징벡터, PLPCC 특징벡터, SMFCC 특징벡터, 왜도, 첨도를 나타낸다.

3.5. 스코어 계산

각 화자의 발성에서 추출된 특징 파라미터를 사용하여 미리 훈련된 GMM에 대하여 임의의 화자의 발성으로부터 추출된 특징 파라미터의 스코어는 식 (11)과 같이 계산된다.

$$Score(i) = \sum_{t=0}^{T-1} \log P(x_t | S_i) \quad (11)$$

여기서, x_t 는 t 번째 프레임에서 추출된 특징 파라미터, S_i 는 i 번째 화자의 모델 파라미터, $P(c_t | S_i)$ 는 i 번째 화자가 c_t 를 발생시킬 확률, T 는 추출된 총 특징벡터의 개수이다.

IV. 실험 및 결과

4.1. 데이터베이스

실험에 사용한 음성 데이터베이스는 미국 국립 표준

기술 연구소에서 구성한 TIMIT 데이터베이스를 사용하였다. TIMIT 데이터베이스는 미국 주요 8개 지역의 남성 438명 여성 192명으로 총 630명의 화자가 2~4초 동안의 각기 다른 10문장을 표준어 또는 방언으로 발성하여 총 6,300개의 음성데이터로 구성되었다. 화자모델 훈련을 위하여 화자 당 임의의 8개의 음성 데이터를 사용하였으며, 모델 훈련에 사용되지 않은 2개의 음성 데이터는 테스트 데이터로 사용되었다.

4.2. 특징 파라미터 및 모델링

특징 추출을 위한 단구간 분석 프레임의 크기는 이전 프레임과 20ms 중첩된 30ms이며, 선 강조 필터 계수는 0.97, DFT 크기는 512이다. 캡스트럼 파라미터 추출을 위한 인지적 필터뱅크의 총 개수는 33개였다. 이러한 조건 하에서 각 단구간 분석 프레임에서는 에너지 프레임에서 추출된 110차의 특징 파라미터는 최종적으로 PCA 또는 LDA를 수행하여 기존의 화자인식에 사용되는 MFCC 36차와 동일한 차원을 갖도록 하였다. 각 화자의 확률분포함수는 혼합 2, 4, 8, 16인 GMM으로 모델링하였다.

4.3. Glottal flow의 왜도와 첨도의 분포 특성

2.4장에서 유성음에서 추출한 왜도와 첨도는 화자인식을 위한 특징파라미터로서 유용하게 활용될 수 있다고 설명하였다. glottal flow 또한 유성음에서 추정되며, 화자의 성대의 떨림 특성을 반영하고 있기 때문에 화자인식을 위한 특징파라미터로서 사용될 수 있다고 판단할 수 있다. 본 논문에서는 임의의 화자 발성에서 추정된 glottal flow에서 왜도와 첨도를 추출하여 분포 특성을 분석함으로써 이를 검증하였다. 그림 5의 (a)와 (b)는 각각의 화자에서 추출한 왜도와 첨도의 분포 특성을 나타내었다. 그림 5를 통해 왜도와 첨도는 각각의 화자마다 다른 분포 특성을 갖고 있으므로 화자인식에 사용되는 특징파라미터로서 유용하게 활용될 수 있다는 것을 알 수 있다.

4.4. SMFCC에 의한 화자인식 성능 분석

Glottal flow에 대한 저역통과 필터링의 최적 차단주파수를 선택하기 위하여 차단주파수를 4 kHz에서 7.5 kHz까지 0.5 kHz간격으로 증가시키며 전수조사를 수행하였다. 그림 6은 저역통과 필터링을 수행한 glottal

flow에서 추출한 SMFCC에 대한 화자인식 결과이다. 결과를 통해 6 kHz에서 56.98%로 가장 높은 인식률을 보이는 것을 확인하였고, 이후 실험에서 차단주파수는 6 kHz로 고정하여 특징을 추출하였다. glottal flow에서 기존의 방식으로 추출한 SMFCC와 본 논문에서 제안한 특징 파라미터를 사용한 화자인식률을 그림 7에 나타내었다.

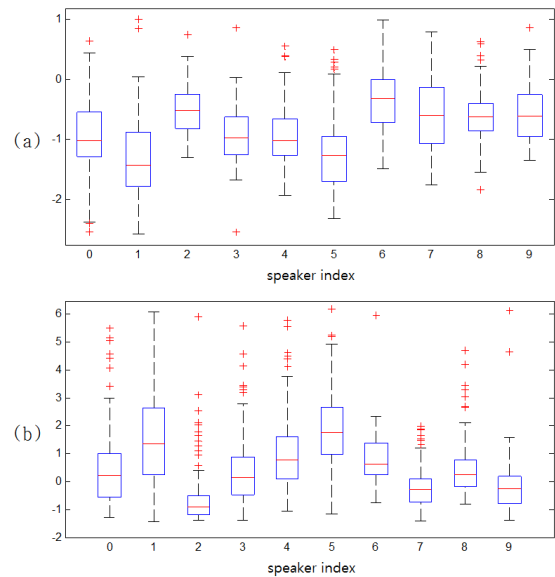


Fig. 5 Distribution characteristics of skewness and kurtosis extracted from glottal flow ((a) distribution of skewness (b) distribution of kurtosis)

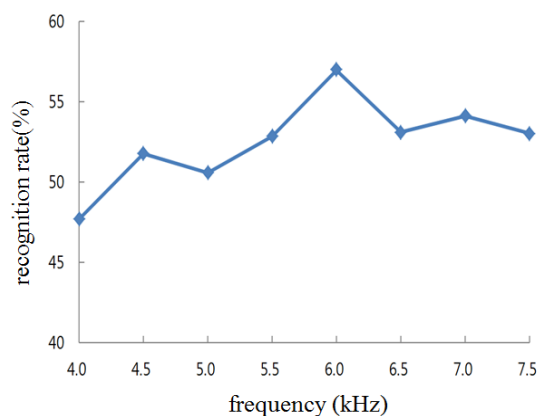


Fig. 6 Recognition results according to the cutoff frequency of lowpass filter (number of Gaussian mixtures : 16)

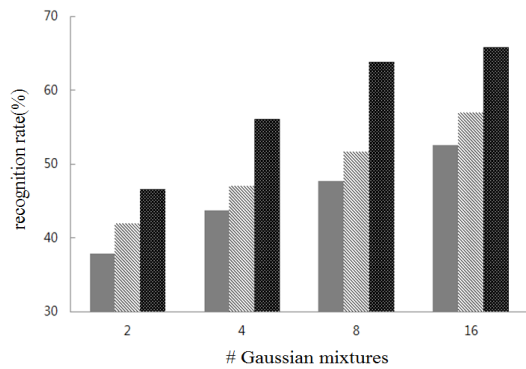


Fig. 7 Recognition results for the proposed feature parameters (■: SMFCC, ▨: SMFCC_L, ■: SMFCC_LSK)

기존의 전 주파수 응답을 활용하는 SMFCC 추출방법보다 저역통과 필터링 및 스펙트럼 확장을 수행한 후 SMFCC를 추출하였을 때(SMFCC_L) 더 좋은 인식률을 보이는 것을 확인 하였다. 또한, glottal flow에서 추출된 왜도와 첨도를 추가로 구성하였을 때(SMFCC_LSK) 가장 좋은 인식률을 보이는 것을 확인하였다. 참고로, 가우시안 혼합 16 에서 SMFCC_L 기준으로 왜도(S) 및 첨도(K)에 의한 성능 향상은 각각 3.20%, 5.18%였다.

4.5. 전체적인 화자인식 성능 비교

MFCC 특징 파라미터만을 활용한 기존의 화자인식 시스템과 본 논문에서 제안된 화자인식 시스템의 화자인식률을 그림 8에 나타내었다.

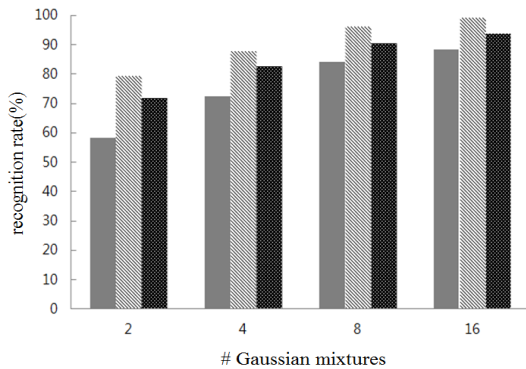


Fig. 8 Comparative speaker recognition results (▨: conventional speaker recognition system, ▨: proposed (PCA), ■: proposed (LDA))

기존 화자인식 시스템과 비교하였을 때 PCA를 활용하여 제안된 화자인식은 가우시안 혼합 개수 2, 4, 8, 16에 대하여 각각 21.03%, 15%, 11.98%, 10.63%의 성능향상을 보였으며, LDA를 활용하여 제안된 화자인식은 가우시안 혼합 개수 2, 4, 8, 16에 대하여 각각 13.57%, 10.19%, 6.34%, 5.39%의 성능향상을 보였다. 결론적으로 본 논문에서 제안한 화자인식 시스템은 기존 화자인식 시스템보다 전반적으로 더 좋은 화자인식 성능을 보였으며, 특히 가우시안 혼합이 낮을 때 더 높은 성능향상을 보였다.

V. 결론

본 논문에서는 특징 파라미터 추출방법을 제안하고 기존의 특징들과의 결합을 통해 화자인식 성능 개선을 도모하였다. 실험결과 제안한 방법으로 추출된 특징 파라미터가 기존의 방법으로 추출된 특징 파라미터보다 높은 인식률을 보이는 것을 확인할 수 있었다. 또한, 기존의 MFCC, PLPCC와 제안된 특징 파라미터를 PCA, LDA를 사용하여 결합하여 화자를 인식하였을 때 기존의 화자인식 시스템보다 전반적으로 우수한 성능개선을 보이는 것을 확인할 수 있었으며, 특히 가우시안 혼합이 낮을 때 높은 성능 향상을 보이는 것을 알 수 있었다. 향후, 제안된 특징 파라미터를 활용하여 deep neural network (DNN) 기반의 화자인식 연구를 계획하고 있다.

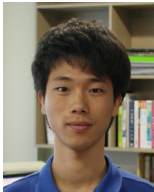
ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2014R1A1A2059191).

REFERENCES

[1] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, Vol. 52, No. 1, pp. 12-40, 2010.

- [2] B. Putra and Suyanto, "Implementation of secure speaker verification at web login page using Mel Frequency Cepstral coefficient-Gaussian Mixture Model (MFCC-GMM)," *ICA*, pp. 358-363, 2011.
- [3] N. Ahmed, "How I came up with the discrete cosine transform," *Digital Signal Processing*, Vol. 1, No. 1, pp. 4-9, 1991.
- [4] D. Reynolds and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech and Audio Proc.*, Vol. 3, No. 1, pp. 72-83, 1995.
- [5] L. Rabiner and B. H. Juang, *Fundamental of Speech Recognition, Signal Processing Series*, Prentice Hall, New Jersey, 1993.
- [6] T. Kinnunen and P. Alku, "On separation glottal source and vocal tract information in telephony speaker verification," *ICASSP*, pp. 4545-4548, 2009.
- [7] J. Markel and A. Gray Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
- [8] P. Alku, H. Tiitinen and R. Näätänen, "A method for generating natural-sounding speech stimuli for cognitive brain research," *CLINPH*, pp. 1329-1333, 1999.
- [9] W. Kleijin and K. Paliwal, *Speech Coding and Synthesis*, 2nd ed., Elsevier, 1998.
- [10] C. Nikias and A. Petropulu, *Higher-Order Spectra Analysis*, Prentice Hall, 1993.
- [11] Martinez, A.M. and Kak, A.C., "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 228-233, 2001.
- [12] Md Jahangir Alam, T. Kinnunen, P. Kenny, P. Ouellet and D. O'Shaughnessy, "Multitaper MFCC and PLP features for speaker verification using i-vectors," *Speech Communication*, Vol. 55, No. 2, pp. 237-251, 2013.
- [13] H. Hermanski, "Perceptually linear predictive(PLP) analysis of speech," *J. Acoust. Soc. Am.*, Vol. 87. No. 4, Apr. 1990.



강지훈(Jihoon Kang)

2013년 경상대학교 전자공학과(학사)
 2015년 경상대학교 전자공학과(석사)
 2015년 ~ 현재 경상대학교 전자공학과(박사과정)
 ※ 관심분야 : 음성신호처리, 화자인식



김영일(Youngil Kim)

1979년 경북대학교 전자공학과(학사)
 1981년 연세대학교 전자공학과(석사)
 1985년 연세대학교 전자공학과(박사)
 1987년 ~ 현재 경상대학교 전자공학과/공학연구원(교수)
 ※ 관심분야 : 음향공학, 음성신호처리



정상배(Sangbae Jeong)

1997년 부산대학교 전자공학과(학사)
 1999년 한국과학기술원 전기 및 전자공학과(석사)
 2002년 한국과학기술원 정보통신공학과(박사)
 2002년 ~ 2006년 삼성종합기술원 컴퓨팅랩(책임연구원)
 2006년 ~ 2009년 한국과학기술원 디지털미디어랩(연구조교수)
 2009년 ~ 현재 경상대학교 전자공학과/공학연구원(부교수)
 ※ 관심분야 : 음성신호처리, 음성오디오 부호화