HOS 특징 벡터를 이용한 장애 음성 분류 성능의 향상

이지연(ICU), 정상배(ICU), 최홍식(연세대), 한민수(ICU)

<차 례>

<Abstract>

Performance Improvement of Classification Between Pathological and Normal Voice Using HOS Parameter

Ji-Yeoun Lee, Sangbae Jeong, Hong-Shik Choi, Minsoo Hahn

This paper proposes a method to improve pathological and normal voice classification performance by combining multiple features such as auditory-based and higher-order features. Their performances are measured by Gaussian mixture models (GMMs) and linear discriminant analysis (LDA). The combination of multiple features proposed by the frame-based LDA method is shown to be an effective method for pathological and normal voice classification, with a 87.0% classification rate. This is a noticeable improvement of 17.72% compared to the MFCC-based GMM algorithm in terms of error reduction.

* Keywords: Pathological voice detection, Gaussian mixture model, Linear discriminant analysis, Classification and regression tree.

1. 서 론

오랫동안, 객관적이고 자동적인 장애 음성 분류를 위해, 음향학적 파라미터에 기반을 둔 많은 연구가 진행되어 왔다. 시간 또는 주파수 영역에 기반을 둔 파라미터 중에서, 장애 음성 판별에 중요한 역할을 하는 특징으로는 pitch, jitter, shimmer, harmonics-to-noise ratio (HNR), normalized noise energy (NNE) 등이 있다. 이 특징 파라미터들은 기본 주파수(fundamental frequency)에 기반을 두고 장애 음성사이에 상관관계가 존재한다고 발표되었으나, 장애 음성에서는 가산 잡음이나 성대의 손상 등으로 인하여 음성이 왜곡되기 때문에 정확한 기본 주파수를 구하기 어렵다 [1].

최근 장애 음성의 판별 연구는 Gaussian mixture model (GMM), neural network (NN), 그리고 vector quantization (VQ)과 같은 패턴 분류 알고리즘을 이용하여 좋은 성능을 보이고 있다 [1]-[4]. 이들 방법 중에서 GMM은 장애 음성 분류에 가장 좋은 성능을 보인다고 발표되었다 [4]. 본 논문에서는, GMM과 linear discriminant analysis (LDA)를 이용하여 널리 사용되는 주파수 공간의 특징 벡터인 mel-frequency cepstral coefficient (MFCC)와 시간 영역 특징 벡터인 higher-order statistics (HOS)의 3차와 4차 통계 변수인 왜곡도(skewness)와 첨도(kurtosis)를 결합하여 성능과 그 변화를 살펴보았다. 실험 결과, 장애 음성과 정상 음성의 분류에 HOS 특징 결합이 좋은 성능을 보임을 알 수 있었으며, 특히 특징 파라미터를 LDA로 축소시킨 후 GMM를 이용하여 변별 결정을 내린 frame HOS-based LDA 방법의 경우 87.0%의 최고 분류 성능을 보였으며, 기존의 MFCC-basd GMM 알고리즘의 성능과 비교하여 에러 제거 측면에서 약 17.72% 개선되었다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문의 연구배경으로써 GMM, HOS, 그리고 LDA 알고리즘에 대해 설명하고, 3장에서는 장애 음성 판별에 사용되는 특징 벡터로써 MFCC, 로그에너지, 왜곡도와 첨도에 대해 살펴보고 장애음성과 정상음성에서 어떤 특징을 보이는 지에 대해 정리한다. 4장에서는 MFCC-based GMM 알고리즘, frame HOS-based LDA 방법, 그리고 sentence HOS-based LDA 방법의 실험과 결과를 분석하고 마지막으로 5장에서 결론을 맺는다.

2. 연구 배경

2.1. Gaussian Mixture Model

GMM은 복수 개의 Gaussian 확률밀도함수로 데이터의 분포를 모델링하는 방법이다. 즉 특정 파라미터의 기대 값이 Gaussian 분포를 가진다고 가정하고 그에 의한 확률 값을 도출하는 것이다. GMM은 평균과 표준편차만으로 mixture의 차원만큼의 공간에 값들에 대한 특징을 모델링할 수 있기 때문에 널리 이용되고 있다.

최종적인 전체 확률밀도함수는 M개의 Gaussian 확률밀도함수 (혹은 성분 (component))의 선형 결합으로 식 (1)과 같이 표현된다.

$$p(X|\theta) = \sum_{i=1}^{M} p(X|\omega_i, \theta_i) P(\omega_i) \tag{1}$$

여기서, $p(X|\omega_i, \theta_i)$는 데이터 X에 대하여 ω_i번째 성분 파라미터 θ_i로 이루어진 확률밀도 함수를 의미하며, $P(\omega_i)$는 혼합 가중치(mixture weight)로 각 확률밀도 함수 (혹은 성분)의 상대적인 중요도를 의미한다.

혼합 가중치를 사전 확률과 같은 형태로 α_i라고 두면 식 (2)와 같은 제약 조건에 따른다.

$$0 \leq \alpha_i \leq 1 \text{ 그리고 } \sum_{i=1}^{M} \alpha_i = 1 \tag{2}$$

확률밀도 함수가 가우시안 분포를 따를 경우 θ_i는 식 (3)과 같은 파라미터 집합이 된다.

$$\theta_i = (\mu_1, \mu_2, \cdots, \mu_M, \sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2, \alpha_1, \alpha_2, \cdots, \alpha_M) \tag{3}$$

전체 모델을 이루는 각 가우시안 성분은 완전(full), 대각(diagonal), 혹은 정방형 (spherical) 공분산 행렬의 형태를 가질 수 있다. 또한 혼합 성분의 개수는 학습 데이터 집합의 크기에 따라 조절할 수 있다. 따라서 GMM으로 데이터의 분포를 모델링할 경우에 혼합 성분의 개수가 충분히 주어지고, 적절한 파라미터 값들만 주어진다며, 이론적으로는 어떠한 연속적인 분포도 거의 완벽하게 추정하여 모델링할 수 있다 [5].

GMM 학습이란 표본 데이터 집합 $X = \{X_1, X_2, \cdots, X_N\}$가 주어질 경우에 데이터의 로그-우도(log-likelihood)를 최대로 하는 각 혼합 성분 가우시안들의 파라미터들을 추정하여 구하는 문제를 말한다. 일반적으로 GMM은 k-means 알고리즘과 같은 expectation-maximization (EM) 알고리즘으로 최적 모델을 추정하여 결정한다 [5].

Godino *et al.*은 장애 음성을 분류하기 위하여 GMM을 이용하여 여러 논문들을 발표했다. 다소 적은 데이터와 특정 병명에 한정되었지만, 교차 검증 실험 (cross-validation scheme)에 의해 최고 성능 94.07%를 얻었다. 지금까지 장애 음성 분류를 위해 발표된 많은 방법들 중에서 Godino *et al.*에 의한 MFCC-based GMM 알고리즘이 가장 신뢰성 있는 방법에 이라고 의한 실험이라고 언급되고 있다 [3][4].

2.2. Higher-order Statistics

HOS는 신호처리의 새로운 기법으로서 과학 및 공학의 다양한 분야에 응용되고 있다. Cumulant라는 용어로서 잘 알려져 있는 HOS와 이와 관련된 푸리에 변환, 즉 다중 스펙트럼(polyspectra)은 신호 처리 과정에서 신호의 크기 정보(amplitude information) 뿐만 아니라 위상 정보(phase information)를 제공하여, 위상 정보를 나타내지 않는 2차 통계적 특성, 즉 연관 특성(correlation)에 비해서 우월한 장점으로 부각된다. 일반적으로 신호처리 과정에 있어서 다중 스펙트럼 분석을 사용한 동기는 세 가지 측면에서 고찰할 수 있다. 첫 번째로, 다중 스펙트럼을 이용하여 검파(detection), 계수 예측(parameter estimation) 및 분류(classification)과정에 있어서 미지의 스펙트럼 특성(unknown spectral characteristics)을 갖고 있는 Gaussian 잡음을 축소시키는 것이다. 이러한 점은 관측된 데이터의 cumulant 스펙트럼으로부터 신호의 계수(parameter)를 검출 혹은 추정하는데 큰 장점으로 인식되고 있다. 두 번째는 신호의 크기와 위상 응답특성 및 시스템을 규명할 수 있는 능력이다. 이러한 동기부여는 다중스펙트럼이 신호의 위상특성을 그대로 유지할 수 있다는 사실에 근거를 두고 있다. 그리고 이러한 특성은 비최소위상(non-minimum phase) 신호 및 시스템을 규명하는 데 중요한 역할을 하고 있다. 세 번째는 time series에 있어서의 비선형성을 찾아내고 특성을 재현시킬 수 있다는 점이다 [6]. 본 논문에서는 HOS의 첫 번째 특성을 이용하여 장애 음성과 정상 음성을 분류하고자 한다.

유성음의 음성 신호 $x(k)$는 식 (4)와 같이 표현된다 [6].

$$x(k) = s(k) + w(k) \tag{4}$$

여기서 $s(k)$는 성대 떨림에 의해 생성된 non-Gaussian 신호이고, $w(k)$는 장애 음성에서 무시할 수 없는 가산 잡음이다.

장애 음성의 $s(k)$는 불안정한 피치 변화에 의해 특징지어진다. 그것은 성대 움직임이 균형적이지 않고, 성대의 불완전한 폐쇄 때문이다. 또한 장애 음성은 성도(vocal tract)에서의 급격한 기류 변화 때문에 고 주파수 노이즈의 증가 현상을 보인다 [4]. 그 고 주파수 노이즈에 의해 일으켜진 장애 음성의 거친 정도는 $w(k)$에 의해 모델링된다. 그와는 반대로, 정상음성의 $s(k)$는 주기적이고 안정적인 양상을 띤다. 장애 음성 파형과 비교하여, 피치에 관계된 통계치를 정확하게 추측할 수 있다. 결국, $s(k)$로부터 추출된 파라미터의 변화율 차이가 장애 음성과 정상 음성을 분류하기 위한 중요한 키가 될 수 있다 [6].

HOS 방법은 가우시안 노이즈를 약화시키고 non-Gaussian 정보들의 일부를 보존하는 방법으로 널리 알려져 있다. 즉 HOS 분석은 랜덤 과정에서 다소 성공적으로 non-Gaussian 통계치를 추측할 수 있다 [6]. 일반적으로, 랜덤 잡음 $w(k)$는 Gaussian 분포로 모델링될 수 있는 반면, 본 논문에서 사용된 모음 /ah/에서 생성된 $s(k)$는

non-Gaussian으로 모델링될 수 있다. 따라서 HOS 분석이 장애 음성에 응용될 때, 그것은 Gaussian 잡음을 약화시킬 수 있으므로 불안정적이고 불연속적인 $x(k)$ 요소들이 쉽게 추측될 수 있다. 따라서 HOS 방법은 정상 음성과 장애음성 사이에 구별적인 모델링을 위한 중요한 방법일 수 있다 [6][7].

다양한 HOS 통계치 중에서, 3차와 4차 cumulant인 왜곡도(skewness), γ_3와 첨도(kurtosis), γ_4가 특징 파라미터로써 널리 사용되고 있다.

2.3. Linear Discriminant Analysis

본 논문에서, 장애 음성과 정상 음성의 분류 성능을 개선하기 위해 중요한 점은 장애 음성과 정상 음성 클래스를 더욱 더 쉽게 구별할 수 있는 다양한 파라미터를 새로운 파라미터 공간 안으로 어떻게 변환할 것인가이다. LDA는 클래스간 분산(between-class scatter)과 클래스내 분산(within-class scatter)의 비율을 최대화하는 주축으로 사상시켜 선형 부공간으로 데이터에 대한 특징 벡터의 차원을 축소하는 방법이다. 즉 LDA는 데이터의 최적 분류의 견지에서 가능한 클래스간의 분별 정보를 최대한 유지시키면서 데이터를 축소하는 방법이라고 할 수 있다 [2][5][8]. Fisher는 클래스내 분산이라는 척도로 평균간의 차이를 정규화하여 함수로 표현하고, 사영된 데이터들의 중심 (평균)간의 거리를 최대화하는 변환행렬을 찾아내는 방법을 제안하였다. 결국 동일한 클래스의 표본들은 인접하게 사영이 취해지고, 동시에 클래스간의 사영은 중심이 가능한 멀리 떨어지게 하는 변환행렬을 찾아내는 것이다. 이것을 Fisher's linear discriminant라고 한다 [5]. 따라서 LDA가 장애 음성과 정상 음성의 분류를 위해 전 처리 과정으로 응용될 때, 그 음성들을 구분하기 위한 최적의 변환 행렬을 구현하는 것이 가능하다 [8].

3. 장애 음성 판별에 사용되는 특징 벡터

3.1. MFCC와 로그 에너지

MFCC는 사람의 귀가 주파수 변화에 반응하게 되는 양상이 선형적이지 않고 로그스케일과 비슷한 멜(Mel) 스케일을 따르는 청각적 특성을 반영한 켑스트럼 계수 추출 방법이다. 그리고 음성 인식의 기본적인 정적 특징 파라미터에 시간의 개념을 부가하면 인식율의 향상을 가져올 수 있다. 그러므로 음성 신호에서 시간 정보를 추출해내기 위해서는 첫 번째 계수와 두 번째 계수의 차로부터 시간의 차이를 얻을 수 있다 [9].

본 논문에서는 fast Fourier transform (FFT)을 이용하여 주파수영역에서 각 멜-스케일로 이루어진 필터뱅크의 출력 값들을 로그변환을 수행한 후 역퓨리에 변환을 통하여 얻

어지는 12차 MFCC , 1차 정규화된 로그 에너지, 그리고 1차 미분 에너지를 GMM 훈련을 위한 파라미터의 초기 입력으로 사용한다.

Saenz-Lecho *et al.*에 의한 가장 신뢰성 있는 방법이라고 의한 언급된 바 있는 Godino *et al.*의 연구는 장애음성에서 MFCC의 특징 양상을 조사하고, 그것을 이용하여 장애음성과 정상음성의 분류 성능을 개선하였다 [4].

3.2. 왜곡도와 첨도

왜곡도, γ_3란 평균 주변의 3차 적률(moment)을 표준편차로 정규화한 것으로 평균에 대한 분포의 비대칭 정도를 나타내는 지표이다. 분포가 좌우대칭일 때 왜곡도는 0이 되고, 왜곡도가 양수이면 분포의 비대칭 꼬리가 양의 값 쪽으로 (skewed right), 왜곡도가 음수이면 분포의 비대칭 꼬리가 음의 값 쪽으로 (skewed left) 치우친다. 첨도, γ_4란 왜곡도와 함께 정규분포로부터 얼마만큼 일탈해 있는가를 나타내는 지표이다. 분포의 뾰족한 정도를 정규분포와 비교하여 나타내는 것으로 양의 첨도는 상대적으로 더 뾰족하고 (γ_4>3, leptokurtic), 음의 첨도는 덜 뾰족한 것 (γ_4 <3, platykurtic)을 나타낸다 [6].

왜곡도와 첨도는 식 (5)와 같이 정의된다 [6].

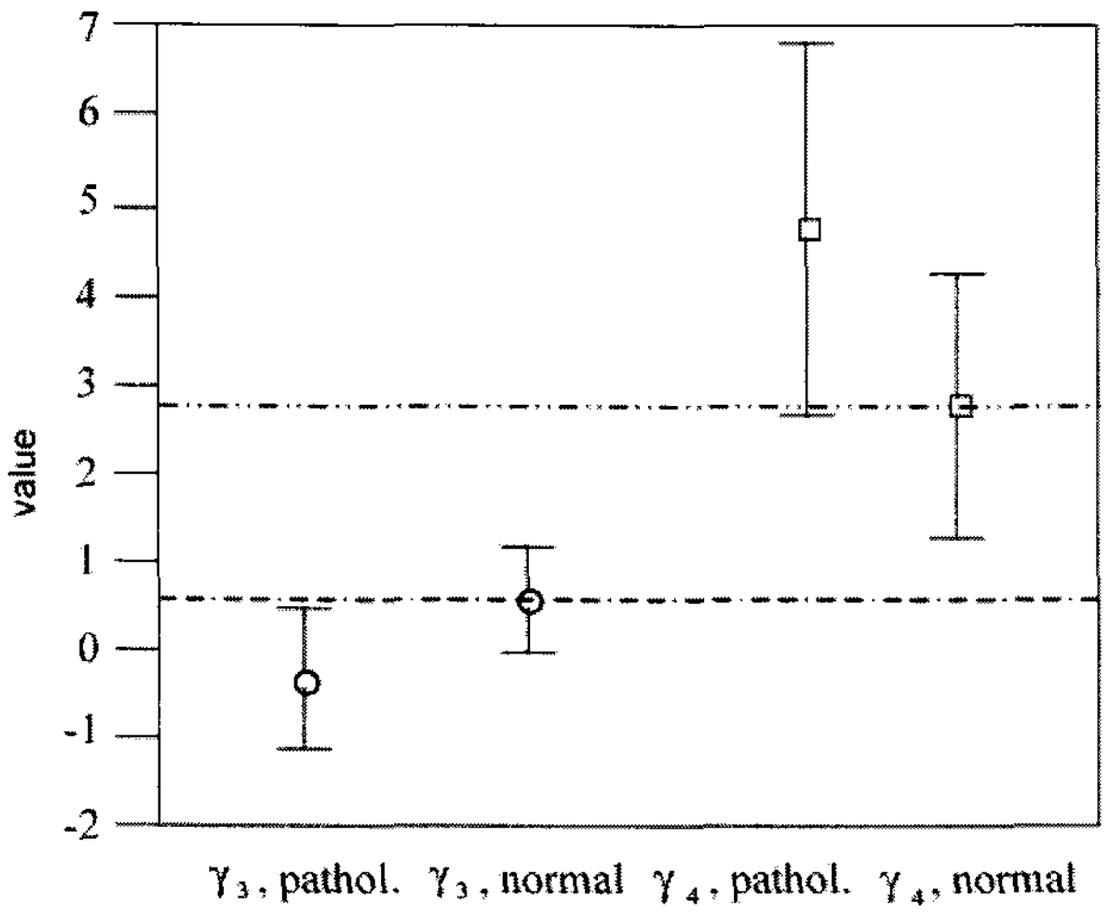
$$\gamma_3 = \frac{\sum_{n=1}^{N}(x_n-\mu)^3}{(N-1)\sigma^3}, \quad \gamma_4 = \frac{\sum_{n=1}^{N}(x_n-\mu)^4}{(N-1)\sigma^4} \tag{5}$$

여기서 x_n은 n번째 음성 샘플, 그리고 N은 전체 샘플 수를 가리킨다. μ와 σ는 각각 x_n의 평균과 표준 편차를 나타낸다.

<그림 1>은 Kay Elemetrics에서 배포한 장애 음성과 정상 음성에서 추출된 γ_3와 γ_4의 분포를 보인다. γ_3의 분포에서, 장애 음성은 skewed left 한 경향을 띠고, 정상 음성은 skewed right한 특징을 보인다. γ_4의 분포에서, 장애 음성은 leptokurtic 분포 (γ_4> 3)를, 정상 음성은 platykurtic 분포 (γ_4< 3)를 가진다. 그리고 장애 음성은 명백하게 γ_3와 γ_4 분포에서 정상 음성보다 큰 변화율을 보인다.

4. 실험 및 결과

Kay Elemetrics에 의해 배포된 장애 음성 데이터베이스(53명의 정상과 600명의 장애 음성)가 본 실험에서 이용되었다 [10]. 화자 수의 균형을 맞추기 위해, 547명의 한국인 정상 음성이 음성 전문가 그룹의 세밀한 조사 후에 첨가되었다. 따라서



<그림 1> γ_3와 γ_4의 분포들 (○: γ_3의 평균, □: γ_4의 평균, 수직 선: γ_3와 γ_4 표준 편차들)

각 600명에 의해 한번 씩 발성된, 즉 각 클래스당 600개의 /ah/ (1~3초) 음성 데이터가 사용되었다. 이때 한국 정상인 음성의 녹음 환경은 Kay Elemetrics에 의해 배포된 데이터베이스의 환경과 비슷하다. 각 데이터는 16 kHz로 다운샘플되었고, 데이터의 70%와 30%가 훈련과 테스트를 위해 각각 사용되었다. 또한 성능 평가를 위해 30 번의 교차 검증 실험(30-fold cross-validation)을 실행했다.

4.1. MFCC-based GMM 알고리즘

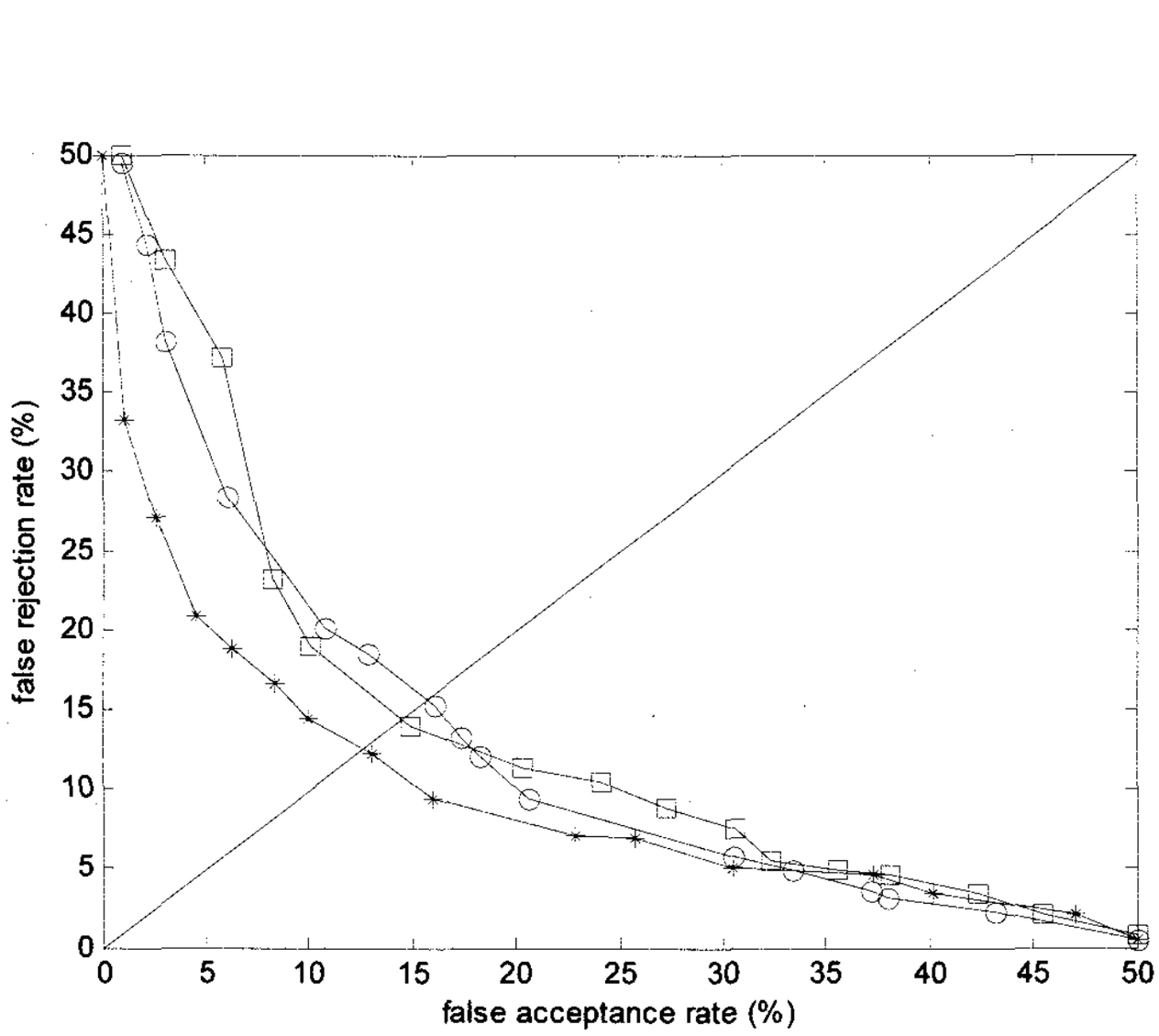
GMM은 Linde-Buso-Gray (LBG) 알고리즘으로 초기 파라미터들을 추측한 후에 EM 알고리즘에 의해 2, 4, 8, 16, 그리고 32 mixtures로 훈련하였다. 그리고 12차 MFCC와 1차 정규화된 로그 에너지 또는 1차 델타 에너지를 특징 파라미터로 하여 13차 또는 14차 벡터를 GMM 파라미터의 초기화 입력으로 사용하였다. 본 논문에서 1~3초 가량 발성된 /ah/ 모음을 데이터베이스로 사용했기 때문에, 성능 면에서, MFCC의 미분값은 MFCC와 비교하여 큰 분별차이가 없다. 따라서 오직 정적 벡터들만을 사용하였다 [3].

<표 1>은 Gaussian mixture의 수와 멜 주파수 기반 필터 뱅크 에너지에 따른 MFCC-based GMM 알고리즘의 평균 equal error rate (EER)와 95% confidence interval (CI)을 보여준다. 95% CI의 정의는 논문 [3]에서 찾을 수 있다. CI는 평균 EER이 얼마나 신뢰성이 있는지를 묘사하는 지수이다. <표 1>에서, 비록 필터 뱅크 차수 축소에 따른 성능은 비슷하지만, mixture 개수가 클수록 성능이 개선됨을 보인다. 또한 13차 보다는 14차가 전반적으로 더 좋은 성능을 보인다. 그와는 반대로, 멜 주파수 기반 필터 뱅크 에너지 수의 증가는 성능에 거의 영향이 없다. Gaussian mixture의 수가 16이고 멜 주파수 기

반 필터 뱅크 에너지가 DCT에 의해 34에서 14로 줄여질 때, 가장 좋은 성능, 즉 평균 EER은 15.8%이다. 이때의 receiver operating characteristic (ROC) 곡선은 <그림 2>에서 찾을 수 있다.

<표 1> MFCC 기반 GMM 알고리즘의 평균 EER±CI (%)

		Mix. 2	Mix. 4	Mix. 8	Mix. 16	Mix. 32
Filterbank 22	13차	24.0±2.1	21.3±2.0	19.9±2.3	17.8±2.0	18.0±1.8
	14차	23.7±2.0	20.1±2.2	19.1±2.2	16.9±1.9	17.6±1.6
Filterbank 26	13차	22.8±1.6	20.0±1.1	18.7±1.6	18.2±1.4	19.5±1.5
	14차	22.1±1.5	19.5±1.3	17.9±1.3	17.6±1.2	18.8±1.4
Filterbank 30	13차	20.1±1.4	21.0±1.2	19.3±1.1	18.1±1.2	19.7±1.4
	14차	19.7±1.2	20.6±1.0	18.9±0.7	16.8±0.8	18.8±1.2
Filterbank 34	13차	20.8±1.2	18.5±0.8	18.1±0.9	16.0±0.6	18.0±0.7
	14차	19.5±1.1	17.5±0.8	17.6±0.6	15.8±0.5	17.4±0.8
Filterbank 38	13차	21.0±1.5	19.7±1.3	20.5±1.5	20.0±1.7	19.8±1.3
	14차	20.7±1.4	19.5±1.2	19.9±1.3	19.1±1.7	19.7±1.5
Filterbank 42	13차	20.8±1.5	22.0±1.6	20.1±1.3	19.1±1.8	18.5±1.5
	14차	20.7±1.3	21.7±1.4	19.5±1.1	18.8±1.3	17.8±1.5



<그림 2> ROC 곡선들 (-□-: 문장 기반 LDA, -*- : 프레임 기반 LDA, -○- : MFCC 기반 GMM 알고리즘)

4.2. HOS-based LDA 알고리즘

본 연구에서, 주파수와 시간 영역의 서로 다른 파라미터들이 LDA-based 방법에 의해 구현된다. 즉 두 가지의 LDA-based 방법들이 장애 음성과 정상 음성의 분류 성능을 개선하기 위해 제안된다. 첫 번째는 각 분석 프레임에서 추측된 멜 주파수 기반 필터 뱅크 에너지, γ_3와 γ_4로 이루어진 식 (6)과 같은 벡터를 구성한다. 이것은 frame HOS-based LDA 방법으로 명명한다.

$$\vec{g} = [FBE_1, FBE_2, \cdots, FBE_P, \gamma_3, \gamma_4] \tag{6}$$

여기서 FBE_k는 k번째 멜 주파수 기반 필터 뱅크 에너지, P는 FBE_P의 전체 개수를 가리킨다.

본 실험에서, 최적의 파라미터 차원을 찾기 위해 멜 주파수 기반 필터 뱅크 에너지의 개수는 22부터 42개의 범위에서 사용된다. 그러므로 $\bar{g}$의 차원은 24부터 44개의 범위를 가진다. 이 때 식 (6)은 LDA 변환에 의해 13차와 14차의 파라미터 벡터로 축소된다. 마지막으로, 그 변환된 벡터는 EM 알고리즘에 의해 GMM을 훈련하기 위해 초기 파라미터로 사용된다.

두 번째 방법은 식 (7)과 같이 각 문장에서 추출된 GMM의 log-likelihoods, γ_3와 γ_4를 이용한다. 그리고 식 (7)의 3차 벡터는 LDA변환에 의해 1차 스칼라로 축소된다. 이것은 sentence HOS-based LDA 방법이라고 명명한다. Log-likelihood ratio의 정의는 참고문헌 [3]에서 찾을 수 있다.

$$g^t = [\overline{LL}, \overline{\gamma_3}, \overline{\gamma_4}] \tag{7}$$

여기서 $\overline{LL}$은 MFCC 기반 GMM 알고리즘에서 추출된 평균 log-likelihood, $\overline{\gamma_3}$는 γ_3의 평균, $\overline{\gamma_4}$는 γ_4의 평균을 가리킨다.

GMM과 MFCC 실험을 위해, MFCC-based GMM 알고리즘과 같은 조건이 frame HOS-based LDA 방법에 적용되었다. <표 2>는 frame HOS-based LDA 방법의 성능을 보여준다. 이것은 MFCC-based GMM 알고리즘과 비슷한 특징을 보인다. Gaussian mixture의 수가 16이고 멜 주파수 기반 필터 뱅크 에너지가 LDA에 의해 36에서 14로 축소될 때, 가장 좋은 성능은 평균 EER 13.0%이다. 이때의 ROC 곡선을 <그림 2>에서 찾을 수 있다.

<표 2> Frame HOS-based LDA 방법의 평균 EER±CI (%)

		Mix. 2	Mix. 4	Mix. 8	Mix. 16	Mix. 32
Filterbank 24	13차	19.8±1.6	18.5±1.5	17.6±1.3	15.9±1.4	17.1±1.5
	14차	19.0±1.5	17.8±1.4	17.1±1.4	14.9±1.5	16.7±1.6
Filterbank 28	13차	18.3±1.4	17.5±1.6	17.9±1.5	17.1±1.3	16.0±1.4
	14차	18.0±1.5	16.7±1.4	17.1±1.5	16.3±1.2	15.7±1.3
Filterbank 32	13차	19.7±1.6	18.6±1.4	18.9±1.5	14.5±1.3	16.7±1.5
	14차	19.5±1.5	18.3±1.1	18.2±1.2	13.9±0.9	16.6±1.2
Filterbank 36	13차	18.1±1.6	17.3±1.4	15.0±1.5	13.7±1.4	14.8±1.5
	14차	18.8±1.3	16.7±1.1	14.2±1.2	13.0±0.9	14.4±1.0
Filterbank 40	13차	18.5±1.1	19.0±1.5	16.8±1.4	16.5±1.4	15.7±1.6
	14차	18.4±1.4	18.4±1.2	16.3±1.3	16.0±1.5	15.2±1.5
Filterbank 44	13차	20.2±1.4	19.6±1.3	17.6±1.3	16.3±1.2	17.7±1.5
	14차	19.9±1.1	18.9±1.4	16.9±1.2	15.9±1.3	17.4±1.5

Sentence HOS-based LDA방법에서의 log-likelihoods는 MFCC 기반 GMM 알고리즘에서 가장 좋은 성능을 보이는 조건에서 결정된다. 이때 평균 EER 성능은 14.2%이며 ROC 곡선은 <그림2>에서 보인다. 결론적으로, frame HOS-based LDA방법은 MFCC-based GMM 알고리즘과 sentence HOS-based LDA방법보다 더 좋은 성능을 보인다. 즉 제안된 HOS-based LDA 방법들이 기존의 MFCC-based GMM 알고리즘보다 장애 음성과 정상 음성 분류를 위해 더욱 효과적이라고 말할 수 있다.

<표 3> 여러 가지 방법들의 성능

		평균 성능±CI (%)
베이스라인 알고리즘	MFCC-based GMM 알고리즘	84.2±0.5
제안한 알고리즘	Frame HOS-based LDA 방법	87.0±0.9
	Sentence HOS-based LDA 방법	85.8±0.6

5. 결 론

본 논문에서는 Gaussian mixture model과 linear discriminant analysis를 이용하여 주파수 공간의 특징 벡터인 mel-frequency cepstral coefficient와 시간 영역 특징 벡터인 higher-order statistics의 3차와 4차 통계 변수인 왜곡도(skewness)와 첨도(kurtosis)를 결합하여 성능과 그 변화를 살펴보았다. 파라미터 결합 방법은 MFCC-based GMM 알고리즘, frame HOS-based LDA 방법, sentence HOS-based LDA 방법을 적용하였다. 실험 결과, 장애 음성과 정상 음성의 분류에 HOS 특징 결합이 좋은 성능을 보임을 알 수 있었으며, 특히 특징 파라미터를 LDA로 축소시킨 후 GMM을 이용하여 변별 결정을 내린 frame

HOS-based LDA 방법의 경우 87.0%의 최고 분류 성능을 보였다. 이 방법을 통해 기존의 MFCC-based GMM 알고리즘의 성능과 비교하여 에러 제거 측면에서 약 17.72%가 개선되었다. 지금까지 장애 음성 분류를 위해 발표된 많은 방법들 중에서 Godino *et al.*에 의한 MFCC-based GMM 알고리즘이 가장 신뢰성 있는 방법이라고 언급된 것처럼, 본 논문에서도 GMM은 가장 좋은 변별 성능을 보이는 패턴 인식 알고리즘이라는 것이 증명되었다. 또한 DCT보다는 LDA에 의해서 파라미터를 축소하는 것이 장애 음성과 정상 음성을 분류하는데 더 적합하다는 것을 보인다는 측면에서 본 연구는 중요하다. Frame HOS-based LDA 방법은 지금까지 발표되어왔던 여러 가지 장애 음성 분류 방법들을 보완하는데 유용하게 응용될 수 있다. 앞으로 실제 환경에서 본 연구 결과를 적용하고 응용할 것이고, 장애 음성의 병명을 분류하는 연구를 할 것이다.

참 고 문 헌

[1] D. Michaelis, M. Forhlich, H. W. Strobe, "Selection and combination of acoustic features for the description of pathological voices", *Journal of the Acoustical Society of America*, Vol. 103, No. 3, pp. 1628-1639, 1998.

[2] T. Xiong, V. Cherkassky, "A combined SVM and LDA approach for classification", *Proc. IEEE IJCNN*, Vol. 3, pp. 1455-1459, 2005.

[3] J. I. Godino-Llorente, S. Aguilera-Navarro, P. Gomez-Vilda, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters", *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 10, pp. 1943-1953, 2006.

[4] N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, P. Gomez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection", *Biomedical Signal Processing and Control*, Vol. 1, No. 2, pp. 120-128, 2006.

[5] 한학용, *패턴인식 개론*, 한빛미디어, 2005.

[6] E. Nemer, R. Goubran, S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain", *IEEE Transaction on Speech and Audio Processing*, Vol. 9, No. 3, pp. 217-231, 2000.

[7] J. B. Alonso, J. de Leon, I. Alonso, M. A. Ferrer, "Automatic detection of pathologies in the voice by HOS based parameters", *EURASIP Journal on Applied Signal Processing*, Vol. 2001, No. 4, pp. 275-284, 2001.

[8] 이지연, 정상배, 최홍식, 한민수, "Detection of pathological voice using linear discriminant analysis", *말소리*, 제64호, pp. 77-88, 2007.

[9] 오영환, *음성언어정보처리*, 홍릉과학출판사, 1997.

[10] Kay Elemetrics Corp, *Disordered Voice Database, Ver. 1.03*, 1994.

접수일자: 2008년 2월 12일
게재결정: 2008년 5월 22일

▶ 이지연(Ji-Yeoun Lee) : 교신저자
주소: 305-732 대전광역시 유성구 문지동 103-6 한국정보통신대학교
소속: 한국정보통신대학교(ICU) 음성/음향 정보 연구실
전화: 042) 866-6206
E-mail: jyle278@icu.ac.kr

▶ 정상배(Sangbae Jung)
주소: 305-732 대전광역시 유성구 문지동 103-6 한국정보통신대학교
소속: 한국정보통신대학교(ICU) 음성/음향 정보 연구실
전화: 042) 866-6296
E-mail: sangbae@icu.ac.kr

▶ 최홍식(Hong-Shik Choi)
주소: 135-270 서울시 강남구 도곡동 146-92 연세대학교
소속: 의과대학 영동세브란스병원 이비인후과
전화: 042) 866-6206
E-mail: hschoi@yumc.yonsei.ac.kr

▶ 한민수(Minsoo Hahn)
주소: 305-732 대전광역시 유성구 문지동 103-6 한국정보통신대학교
소속: 한국정보통신대학교(ICU) 음성/음향 정보 연구실
전화: 042) 866-6123
E-mail: mshahn@icu.ac.kr