피처벡터 축소방법에 기반한 장애음성 분류

이지연* 정상배* 최홍식** 한민수*
* 한국정보통신대학교 음성음향정보연구실
** 연세대학교 의과대학 영동세브란스병원 이비인후과학교실 음성언어의학연구소

Classification of pathological and normal voice based on dimension reduction of feature vectors

Ji-Yeoun Lee* SangBae Jeong* Hong-Shik Choi** Minsoo Hahn*
* Speech and Audio Information Lab., Information and Communication Univ.
** Institute of Logopedics and Phoniatrics, Department of Otorhinolaryngology,
Yongdong Severance Hospital, Yonsei University College of Medicine

jyle278@icu.ac.kr, sangbae@icu.ac.kr, hschoi@yumc.yonsei.ac.kr, mshahn@icu.ac.kr

Abstract

This paper suggests a method to improve the performance of the pathological/normal voice classification. The effectiveness of the mel frequency-based filter bank energies using the fisher discriminant ratio (FDR) is analyzed. And mel frequency cepstrum coefficients (MFCCs) and the feature vectors through the linear discriminant analysis (LDA) transformation of the filter bank energies (FBE) are implemented. This paper shows that the FBE LDA-based GMM is more distinct method for the pathological/normal voice classification than the MFCC-based GMM.

I. 서론

많은 연구가 장애음성의 객관적인 분류를 위한 음향학적 파라미터의 추출에 초점이 맞추어져 왔다. 음향학적 파라미터 중에서, pitch, jitter, shimmer, harmonics-to-noise ratio (HNR), normalized noise energy (NNE)들이 특히 중요하다. 이 파라미터들은 fundamental frequency에 기반을 두고 있다[1]. 그러나 장애음성에서 fundamental frequency를 정확하게 추출하기가 쉽지 않다.

최근에, Gaussian mixture models (GMMs), neural networks (NN), 그리고 vector quantization (VQ)와 같은 패턴 인식 알고리즘들의 성능이 장애음성 분류를 위해 보고되고 있다. 특히, 가우시안 믹스처 모델링과 mel-frequency cesptral coefficients (MFCCs)의 특징 벡터의 결합은 불규칙적인 음성 파형의 검출을 위해 더욱 더 유용하다고 보고되어졌다[2][3].

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 장애음성 분류의 연구동향에 대해 언급한다. 3장은 linear discriminant analysis (LDA)에 대해 기술하고, 4장에서는 멜 주파수 기반의 필터 뱅크 에너지의 유용성을 분석한다. 5장에서 Baseline 알고리즘을 기반으로 LDA 변환의 실험 및 결과를 검토한 후, 6장에서 결론을 맺는다.

II. 연구동향

2002년에 Dibazar는 Kay Elemetrics에서 만든 장애음성 데이터베이스를 가지고 장애음성 분류 성능에 가장 좋은 결과를 보였다. 그들은 multi-dimensional voice program (MDVP)에서 추출된 파라미터들, MFCCs, 그리고 fundamental frequency를 사용했다. 그리고 Hidden markov model (HMM)을 이용하여 성능을 측정했다[4][5]. 그리고, 2002년, Hadjitodorov는

jitter, shimmer 그리고 노이즈 측정에 관련된 파라미터를 사용하여 장애음성의 자동적인 분류를 위한 시스템을 제안했다. 그들은 Kay Elemetrics에서 분포된 장애음성 데이터베이스를 사용했다. 그리고 LDA와 nearest neighbour (NN) 클러스터링을 사용하여 장애/정상음성을 분류했다[4][6]. 2006년에 Godino는 장애음성을 분류하기 위해 GMMs과 MFCCs을 이용하였다. 그들은 Kay Elemetrics에서 분포된 장애음성 데이터베이스 중에서 랜덤으로 음성 파일을 선택했다. 피처로써 MFCCs와 그들의 미분 값이 이용되었으며 GMMs을 통해 성능이 측정되었다[3][4].

Pathological voice의 주된 현상은 성대의 무게 증가에 따른 불완전한 패쇄, 비 대칭적인 움직임 그리고 성대 조직 특성의 변화 등에 기인한다. 즉, source에 일어나는 문제점이 음성의 변화를 일으킨다. 장애음성의 특징을 잘 나타내는 파라미터로써, MFCCs가 알려져있다. 장애음성에서, 질량 증가에 따른 mucosal 파형과 관계된 영향은 MFCCs의 low 밴드에 반영되고, 반면에 high 밴드는 개폐 불안정에 따른 노이즈 성분을 모델링 할 수 있다. 그러므로, MFCCs는 speech pathology를 모델링하는데 유용하며, 장애음성 디텍션을 위해 적절하다고 생각된다[3]. 본 논문에서, GMMs는 장애/정상 음성에서 추출된 MFCCs의 분포를 모델링하기 위해 사용된다. 즉 GMMs은 장애/정상 음성을 구별할 수 있는 성도와 후두의 수학적인 모델을 마련한다.

III. Linear discriminant analysis

LDA (선형판별분석)는 principle component analysis (PCA, 주성분분석법)과 더불어 대표적인 특징 벡터 차원 축소 기법 중의 하나이다. LDA는 클래스간 분산 (between-class scatter)과 클래스내 분산 (within-class scatter)의 비율을 최대화하는 방식으로 데이터에 대한 특징 벡터의 차원을 축소하는 방법이다. 즉, LDA는 수식(1)과 같이 $tr(W^{-1}B)$을 최대화하는 선형 변환행렬을 찾는 것이다. 결국, 동일한 클래스의 표본들은 인접하게 사영이 취해지고, 동시에 클래스간의 사영은 중심이 가능한 멀리 떨어지게 하는 변환 행렬을 찾는 것이다. 그리고 그 선형 변환행렬은 가장 큰 고유 값에 관련된 고유벡터로 구성된다[7].

$$\mathbf{W} = \frac{1}{N}\sum_{k=1}^{K}\sum_{n=1}^{n_k}(x_{kn}-\mu_k)(x_{kn}-\mu_k)^t \qquad (1)$$

$$\mathbf{B} = \frac{1}{N}\sum_{k=1}^{K} n_k(\mu_k-\mu)(\mu_k-\mu)^t$$

W : within-class covariance 행렬

B : between-class covariance 행렬

N : 전체 훈련 프레임 개수, K : 클래스 개수

n_k : k^{th}클래스의 훈련 프레임 개수,

μ_k : k^{th}클래스의 평균, μ : 전체 평균

IV. 멜 주파수 기반의 필터 뱅크 에너지의 유용성 분석

3장은 fisher discriminant ratio (FDR)를 이용하여 멜 주파수 기반의 필터 뱅크 에너지의 유용성을 증명한다. FDR은 클래스 또는 피처들이 얼마나 잘 분리되었는가, 그리고, 음성 인식에서 여러 가지 피처들 중에서 가장 변별적이고 효과적인 피처를 선택하고자 할 때 응용할 수 있다. 즉 FDR은 분별력 또는 변별력을 나타내는 척도, 표준으로 널리 사용되고 있다[3]. FDR은 수식(2)처럼 정의된다.

$$F_i = \frac{(\mu_{iC}-\mu_{i\bar{C}})^2}{\sigma_{iC}^{\ 2}+\sigma_{i\bar{C}}^{\ 2}} \qquad (2)$$

μ: 클래스의 평균, σ^2: 클래스의 분산

C : 정상음성의 클래스, $\bar{C}$: 장애음성의 클래스

F_i의 값이 크면 클수록, 그 피처는 여러 가지 클래스들을 구분하는데 더욱 더 변별적이고 중요한 피처임을 나타낸다. 그림 1은 34차 멜 주파수 기반의 필터 뱅크 에너지의 정규화된 FDR을 보여준다. 장애/정상음성을 구별하기 위한 멜 주파수 기반의 필터 뱅크 에너지의 유용성은 low와 high 밴드에서 발견된다. 가장 큰 값은 700 Hz이하의 low 밴드에서 발견되고, 이 부분에는 1st 포만트가 존재한다. 그것은 1st 포만트가 정상음성으로부터 장애음성을 분류하는데 중요한 파라미터가 될 수 있음을 보여준다. 또한 5 kHz 이상의 높은 주파수영역에서 우리는 큰 값들을 볼 수 있다. 이것은, 성대의 불규칙적인 움직임으로 인한 high 주파수 밴드에서의 노이즈 증가으로 짐작할 수 있다. 따라서 우리는 1st 포만트와 high 주파수가 장애/정상음성을 구분하는 중요한 정보임을 알 수 있다.

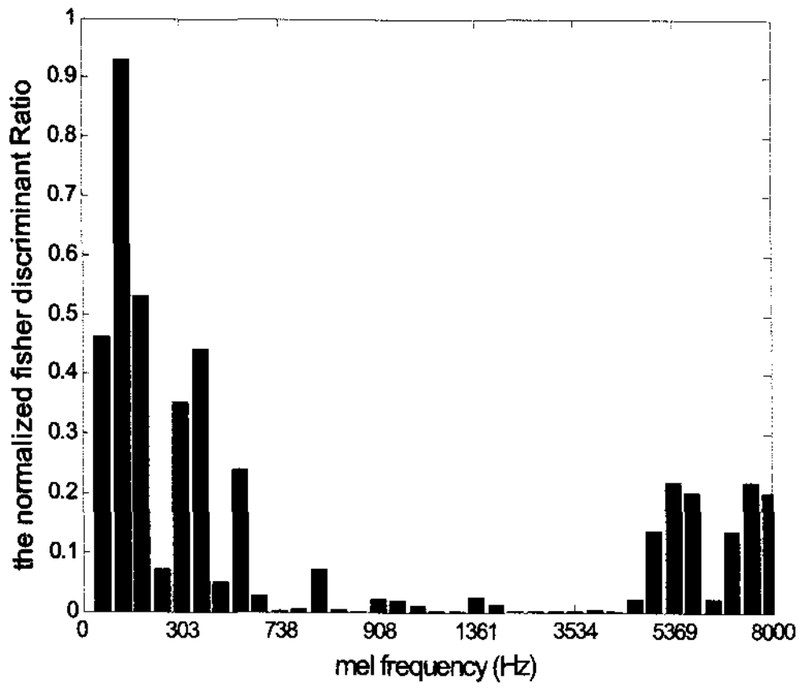

그림 1. 34차 필터 뱅크 에너지들의 정규화 FDR

V. 실험 및 결과

4.1 Database

Kay Elemetrics에서 배포한 장애음성 데이터베이스가 실험에 사용되었다[8]. 본 논문에서는, 위의 데이터베이스로부터, /ah/ 모음으로 구성된 53명의 정상음성과 600명의 장애음성을 이용했다. 그리고 GMMs 훈련을 위해, 547명의 정상 한국인에 의해 발성된 /ah/ 모음이 부가되었다. 녹음 환경은 Kay Elemetrics에서 만든 장애음성 데이터베이스의 환경과 비슷하다. 전체 음성들은 16 kHz로 다운 샘플링되었다. 정상음성과 장애음성 화자의 70%와 30%가 각각 훈련과 테스트 실험을 위해 사용되었다.

4.2 전반적인 블록 다이어그램

그림 2는 장애/정상음성 분류 알고리즘의 전반적인 블록 다이어그램을 보인다. 우선, 멜 주파수에 기반한 필터 뱅크 에너지들이 음성 샘플에서 추출된다. 그것들은 두 가지 변환에 의해 피처 파라미터로 구현된다. 하나는 discrete cosine transformation (DCT)를 통한 MFCCs 추출이고 또 다른 방법은 LDA 변환을 통한 피처 추출이다. 훈련과정에서, 장애/정상음성들의 가우시안 모델 파라미터 (평균, covariance, 믹스처 웨이트)들을 추측하기 위해, expectation-maximization (EM) algorithm을 통해 가우시안 믹스처가 반복적으로 훈련된다. 그리고, 최고 성능을 보일 때의 log-likelihood ratio가 문턱값, θ, 로 결정된다. 테스트 과정에서 계산된 log-likelihood ratio, $\Lambda(X)$, 는 문턱값, θ, 와 비교하여, $\Lambda(X) > \theta$이면 정상음성으로 분류하고, 그 반대면 장애음성으로 분류된다.

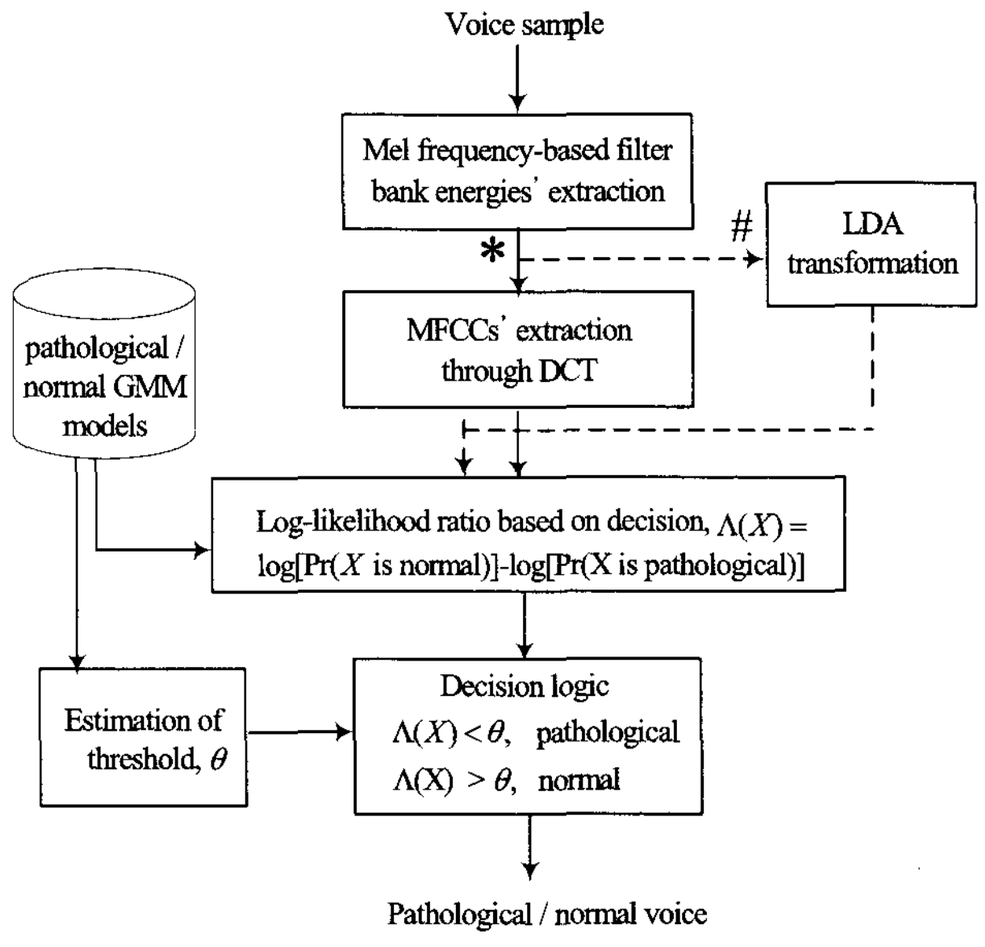

그림 2. 전반적인 블록 다이어그램

4.3 Baseline 성능 (MFCC-based GMM)

GMMs은 Linde-Buso-Gray (LBG) 알고리즘에 의해 초기화된 후, 2, 4, 8, 16, 그리고 32개의 믹스처로 각각 훈련된다. 그리고 22차에서 42차까지의 필터 뱅크 에너지들이 DCT에 의해 MFCCs 12차로 축소된다. MFCCs, 그들의 미분 값과 에너지를 이용하는 것과 MFCCs만을 이용하여 성능을 비교했을 때, 큰 변별적인 차이를 보이지 않는다고 보고된 바 있다[3]. 그러므로, 본 논문에서는 MFCCs 만 이용했다. 표 1은 가우시안 믹스처와 필터 뱅크 개수에 따른 MFCCs에 기반한 GMM 알고리즘의 성능을 보인다. 가우시안 믹스처가 16개이고 34개의 필터 뱅크 에너지가 12개로 축소될 때, equal error rate (EER)은 17%이다. 비록, 차원 축소에 따른 성능들은 비슷하지만, 많은 믹스처로 훈련할 수록 성능을 개선된다고 말할 수 있다. 반대로, 필터 뱅크의 개수는 성능 개선에 큰 효과는 보이지 않는다.

표 1. MFCC-based GMM의 EER (%)

	Mixture 2	Mixture 4	Mixture 8	Mixture 16	Mixture 32
Filter bank 22th	25.00	22.00	20.00	18.00	18.00
Filter bank 26th	22.00	20.00	19.00	19.00	20.00
Filter bank 30th	21.00	21.00	20.00	18.00	20.00
Filter bank 34th	21.00	20.00	19.00	17.00	18.00
Filter bank 38th	21.00	21.00	21.00	20.00	21.00
Filter bank 42th	21.00	22.00	20.00	20.00	19.00

4.4 FBE LDA-based GMM의 성능

훈련한 GMMs의 조건은 MFCC-based GMM에서 사용된 것과 같다. 이때, MFCC-based GMM에서 좋은 성능을 보인 30차와 34차의 필터 뱅크 에너지에서 차원을 축소하고 성능을 측정했다. 표 2는 30차의 필터 뱅크 에너지가 LDA 변환을 통해, 12차, 18차, 그리고 24차로 축소했을 때, EER 성능을 보인다. 전반적으로, MFCC-based GMM 알고리즘보다 더 좋은 성능을 보인다. 성능은 믹스처 개수에 따라 증가하는 경향을 보이며 가장 좋은 성능은 EER 17%이다. 표 3은 34차의 필터 뱅크 에너지가 LDA 변환을 통해, 12차, 18차, 24차, 그리고 30차로 축소했을 때, EER 성능을 보인다. 믹스처 개수와 필터 뱅크 개수가 증가할수록 성능이 개선됨을 보인다. 가장 좋은 성능은 34차의 필터 뱅크 에너지가 24차로 축소되고, 믹스처 개수가 16일 때, 85%이다. 결론적으로, 성능은 FBE-LDA 방법을 통해 약 2%의 성능 개선을 보인다.

표 2. 30차 필터 뱅크 에너지를 사용했을 때 EER(%)

	Mixture 2	Mixture 4	Mixture 8	Mixture 16	Mixture 32
12th	20.00	19.50	19.00	17.00	19.00
18th	20.00	18.00	18.00	17.00	18.00
24th	19.00	18.00	17.00	18.00	19.00

표 3. 34차 필터 뱅크 에너지를 사용했을 때 EER(%)

	Mixture 2	Mixture 4	Mixture 8	Mixture 16	Mixture 32
12th	20.00	20.00	19.00	16.00	17.00
18th	20.00	19.00	18.00	17.00	17.00
24th	20.00	19.00	17.00	15.00	16.00
30th	20.00	19.00	18.00	17.00	17.00

VI. 결론

본 논문의 목적은 장애음성 분류를 위해 효과적인 변별적인 피처 파라미터를 마련하기 위해 필터 뱅크 에너지에 DCT와 LDA 변환을 구현하고 성능을 비교하는 것이다. 우리는 FDR을 이용하여 멜 주파수에 기반한 필터 뱅크 에너지를 분석했다. 그리고 DCT와 LDA 축소 변환을 통한 피처벡터들을 가지고 GMM 검출기를 구현했다. 그리고 LDA 방법을 통한 피처 벡터의 구현과 장애음성 검출사이에 강한 상관관계가 있었다. 가장 좋은 성능은 34차의 피처 벡터가 FBE-LDA 방법을 통해 24개로 축소되고, 믹스처가 16으로 훈련될때, 85%이다. 제안된 FBE-LDA 방법은 잘 알려진 MFCC-based GMM 방법보다 전반적으로 더 좋은 성능을 보인다. 성능은 에러 감소 측면에서 11.77%의 개선을 보인다. 이 결과는 Godino at al.[3]의 결과와 비교하여 좋은 성능을 보인다.

우리가 앞으로 할 연구는 실제 생활에서 우리의 알고리즘을 적용하여 분석하는 연구와 병명을 분류하는 연구를 포함한다.

참고문헌

[1] Dirk Michaelis, Matthias Forhlich, and Hans Werner strobe, "Selection and combination of acoustic features for the description of pathological voices," J. Acoust. Soc. Am. 103(3), March 1998

[2] D. A. Reynolds, R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE transaction on speech and audio processing, Vol* 3, pp. 72-83, January 1995.

[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 transaction on biomedical engineering* : accepted for future publication.

[4] N.Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, P. Gomez-Vilda, "Methodological issues in the developmentof automatic systems for voice pathology detection," Biomed. Signal Proc. Cont.1(2),pp. 120-128 (2006)

[5] A.A. Dibazar, S.Narayanan, T.W.Berfer, "Feature analysis for automatic detection of pathological speech," *in Proc. IEEE EMBS/BMES Conf, Vol* 1, pp. 182-183 (2002)

[6] S. Hadjitodorov, P.Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal disease screening," Med.Eng.phys. 24(6) (2002)

[7] S. Olivier, "On the Robustness of Linear Discriminant Analysis as a Preprocessing Step for Noisy Speech Recognition," *Proc. IEEE conference on acoustics, speech, and signal processing, Vol* 1, pp. 125-128, May 1995.

[8] Kay Elemetrics Corp, "Disordered Voice Database", ver.1.03, 1994