켑스트럼 기반의 후두암 감별을 위한 채널보상*

김영국(부산대), 김수미(부산대), 김형순(부산대), 왕수건(부산대),
조철우(창원대), 양병곤(동의대)

<차 례>

<Abstract>

Channel Compensation for Cepstrum-Based Detection of Laryngeal Diseases

Young Kuk Kim, Su Mi Kim, Hyung Soon Kim, Soo-Geun Wang, Cheol-Woo Jo, Byung-Gon Yang

Automatic detection of laryngeal diseases by voice is attractive because of its non-intrusive nature. Cepstrum based approach to detect laryngeal cancer shows reliable performance even when the periodicity of voice signals is severely lost, but it has a drawback that it is not robust to channel mismatch due to different microphone characteristics. In this paper, to deal with mismatched training and test microphone conditions, we investigate channel compensation techniques such as Cepstral Mean Subtraction (CMS) and Pole Filtered CMS (PFCMS). According to our experiments, PFCMS yields better performance than CMS. By using PFCMS, we obtained 12% and 40% error reduction over baseline and CMS, respectively.

* Keywords: laryngeal disease, voice analysis, channel normalization, pole filtering

* 본 논문은 보건복지부 협동기초연구지원 연구개발사업(02-PJ1-PG10-31401-0005) 연구결과의 일부입니다

1. 서 론

후두질환은 환자의 음성 특성에 큰 변화를 가져오며, 일반인보다 거칠고 쉰 목소리가 나는 증상 등이 그 예이다. 환자의 음성 청취는 후두 질환을 검진하는데 중요한 도구가 되며, 최근 신호처리 기술의 발달과 더불어 음성신호의 자동분석에 의한 후두질환 감별에 많은 관심이 모아지고 있다. 음성분석에 의한 후두질환 감별은 감별성능만 어느 이상 보장된다면, 환자의 고통 없이 검사를 신속 간편하게 할 수 있을 뿐만 아니라, 인터넷 등을 통한 원격 검진이 가능하다는 장점을 지닌다.

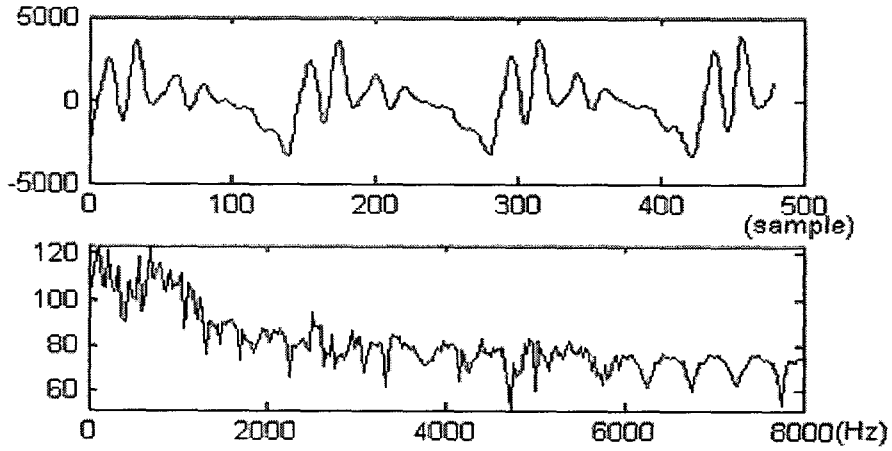
후두질환 감별용 음성 분석방법 중 대표적인 것으로 Multi-Dimensional Voice Program (MDVP) 분석에 의한 jitter나 shimmer파라미터 등을 들 수 있으나[1-6], 이들의 경우 주기성이 크게 훼손된 후두암 말기의 음성 데이터에 대해서는 분석 자체가 불가능해지는 문제점이 있다. MDVP분석이 불가능할 정도로 주기성이 훼손된 음성에 대해 효과적인 감별이 가능한 신뢰도 높은 분석방법으로서, 음성인식 분야에서 널리 사용되어 온 켑스트럼(cepstrum) 파라미터 분석방법을 검토하였다[7]. 그러나 켑스트럼 파라미터의 경우 다른 파라미터보다 마이크 특성 차이 등으로 인한 채널 왜곡에 의한 성능저하가 상대적으로 심하다는 단점이 있다.

본 논문에서는 우선 음성인식 기술을 기반으로 한 후두질환 음성의 자동감별 방식들을 검토하였다. 후두암 음성 감별을 위한 기본적인 특징 파라미터로 음성인식에 널리 사용되는 Linear Predictive Cepstral Coefficients (LPCC)를 사용하고, 인식기로는 Gaussian Mixture Model (GMM) 기반의 분류기를 사용하여, 켑스트럼 차수 및 분석 윈도우 크기를 변화시켜가면서 정상인과 후두암 환자의 음성을 자동 감별하는 실험을 수행하였다. 다음으로 훈련과 테스트 시 마이크 불일치에 따른 성능저하 문제를 해결하기 위한 채널보상 방법에 대한 연구를 하였다. 채널 왜곡으로 인한 불일치를 줄이기 접근 방법은 특징벡터 영역에서의 보상방법과 모델적응 방식으로 나눌 수 있는데, 본 논문에서는 특징벡터 영역에서의 대표적인 보상방법인 Cepstral Mean Subtraction (CMS) 방법을 적용해 보고, 이 방법을 단모음 음성에 적용할 때의 문제점을 극복하기 위해 pole filtered CMS 방법의 적용을 검토하고[8][9] 이들의 성능을 비교평가 하였다.

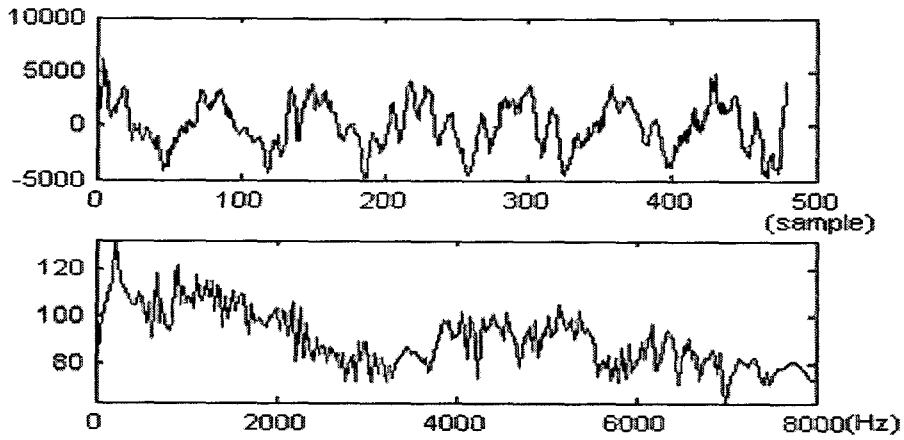
본 논문의 구성은 다음과 같다. 2장에서는 실험에 사용한 후두질환 감별 시스템에 대해 기술하고, 3장에서는 채널보상에 대한 방법, 4장에서 음성 데이터와 실험 방법, 그리고 실험 결과를 다루고, 마지막으로 5장에서 결론을 맺는다.

2. 후두질환 감별 시스템

<그림 1>은 정상인과 후두암 환자가 /아/를 발음했을 때의 음성의 파형과 스펙트럼의 예이다.



(a)정상인



(b) 후두암 환자

<그림 1> 정상인과 후두암 환자의 음성 파형과 스펙트럼의 예(모음 /아/)

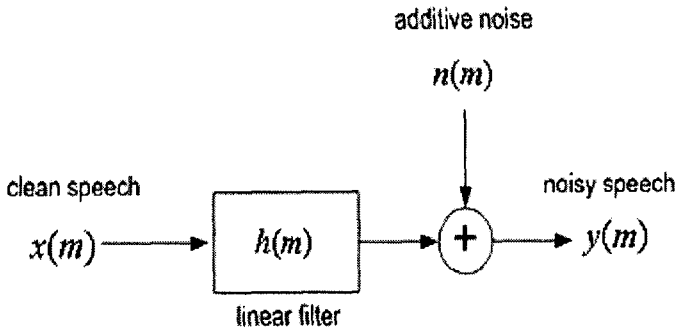
후두질환 환자의 음성은 정상인에 비해 잡음 성분이 많고, 스펙트럼의 고주파 성분이 상대적으로 높으며, 주기성이 떨어지는 성질을 갖는다. 그림으로부터 주기

성 정보 이외에도 음성 스펙트럼의 포락선 정보가 정상인과 후두암 환자의 구별에 효과적으로 사용될 수 있음을 알 수 있다.

본 논문에서 스펙트럼 포락선 정보를 표현하는 특징 파라미터로 LPCC를 사용하였고, 후두질환 감별을 위해 GMM인식기를 사용하였다. GMM을 훈련하기 위해 정상인, 후두암 환자, 양성 환자의 훈련 데이터로 EM알고리즘을 이용해 정상과 악성 그리고 양성의 3개의 모델을 생성하였다. 정상이나 양성인 경우 하나의 그룹으로, 악성 즉 후두암의 경우를 또 하나의 그룹으로 감별하도록 하였다.

3. 채널보상 방법

일반적으로 음성인식 시스템의 성능을 저하하는 다양한 왜곡은 <그림 2>와 같이 크게 시간 영역에서의 convolution 잡음과 부가잡음으로 나눌 수 있다.



<그림 2> 잡음 환경에 대한 왜곡 모델

이러한 왜곡으로 인한 불일치를 줄이기 위해 다양한 접근 방식이 제안되어 왔으며 크게 특징벡터 영역에서의 음질개선(speech enhancement) 방법과 모델적응(model adaptation) 방법으로 나눌 수 있다. 본 논문에서는 마이크 채널 특성 차이에 따른 convolution한 잡음에 대해, 특징벡터 영역에서 CMS와 pole filtered CMS를 통한 채널 보상방법에 대한 실험을 하였다.

3.1. Cepstral Mean Subtraction (CMS)

켑스트럼은 음성인식에서 음성의 특징을 표현하기 위해 가장 널리 사용되는 계수이며, 이에 따라 왜곡에 강인한 특성을 얻기 위한 다양한 정규화 방법이 개발되었다. 이들 중 대표적인 방법이 CMS이다. 이 방법은 시간 영역에서 채널 왜곡

과 같은 convolution 잡음이 켑스트럼 영역에서는 부가적인 형태로 표현되는 것을 이용한다. 즉, 음성으로부터 켑스트럼 계수의 평균을 추정하고, 각 구간 켑스트럼 계수에서 그 평균을 빼주는 방법으로서 구현이 간단하다.

음성 켑스트럼의 장구간 평균이 영이라 가정하면 추정된 켑스트럼은 다음 식으로 표현할 수 있다.

$$C^{t}_{comp} = C^{t}_{y} - m_{y} \tag{1}$$

여기서

$$m_{y} = \frac{1}{T}\sum_{t=1}^{T} C^{t}_{y} \tag{2}$$

이고, C^{t}_{y}와 C^{t}_{comp}는 각각 t번째 프레임에서 보상 전과 후의 켑스트럼 벡터이고, T는 입력음성의 전체 프레임 수를 의미한다.

3.2. Pole Filtered CMS (PFCMS)

음성 스펙트럼에서 국부적인 최대값(local maximum)을 나타내는 포먼트 성분들은 음성 및 화자의 정보를 상당량 포함하고 있다. 이러한 음성 스펙트럼의 특징과는 달리 대부분의 채널 특성은 주파수 영역에서 완만한 굴곡을 가지고 있다. 따라서 LPC 캡스트럼으로 부터 채널 성분 추정 시 캡스트럼 평균을 바로 구하지 않고, 주파수 영역에서의 굴곡을 완만하게 만들어 CMS 과정에서 제거되는 포먼트 성분들을 보존하고자 하는 Pole Filtered CMS (PFCMS) 방법이 제안되었다[8][9]. PFCMS 방법은 채널 성분을 추정할 때 전극(all pole) 모델인 LPC분석을 통해서 얻은 주파수 영역의 전달 함수에서 단위원에 가까운 극점(pole)의 영향을 줄여 채널을 모델링하는 것이다.

3.2.1. Method 1: Conventional PFCMS [8]

<그림 3>(a)와 같이 단위원에 가까이 위치한 극점을 반경이 1 미만인 안쪽 원의 위치로 이동시킴으로써 주요한 극점의 영향이 떨어지게 하면 좁은 대역폭을 갖는 극점의 대역폭이 넓어지게 되며, 그 결과 그림 3(b)와 같이 추정된 채널 스펙트럼의 굴곡이 완만해 지게 된다. Pole-filtered cepstrum을 구하는 과정은 다음과 같다.

1. 매 프레임 t에 대해서 다음의 과정을 수행한다.

1.1. LP polynomial의 근 z_k를 구한다.

1.2. 만약 $abs(z_k) \geq \alpha$를 만족하면 $abs(\tilde{z}_k) = \alpha$를 구한다.

1.3. *pole filtered* $Lpcc(t, n)$을 $\tilde{z}_k$를 이용하여 구한다.

여기서 t와 n은 각각 프레임 번호 및 켑스트럼 차수이다.

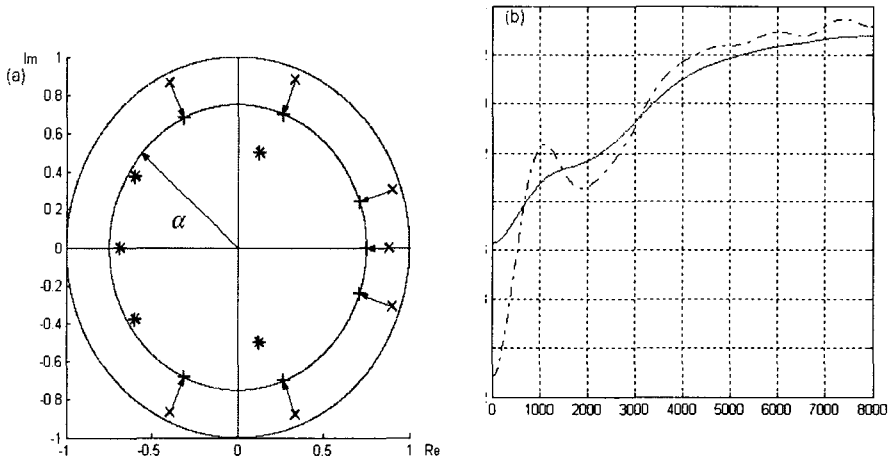
2. 음성 전체 프레임에 대해 채널 성분을 다음과 같이 추정한다.

$$Lpcc_h(n) = \frac{1}{T}\sum_{t=1}^{T} pole\ filtered\ Lpcc(t, n)$$

3. 채널 보상된 켑스트럼을 다음과 같이 구한다.

$$Lpcc(t, n)_{comp} = Lpcc(t, n) - Lpcc_h(n)$$

위 과정에서 $\alpha\ (0 < \alpha < 1)$는 실험에 의해 결정되는 상수 값이다.



(a) pole filtering 방법

(b) pole filtering의 효과
(점선:CMS, 실선: PFCMS)

<그림 3> Pole filtering 방법 및 결과

3.2.2. Method 2: Fast PFCMS [9]

앞에서 설명한 conventioanl PFCMS의 경우 음성 프레임의 LP 다항식의 근을 구해야 하므로 많은 계산량을 필요로 하게 된다. Ramachandran 등은 pole filtering의 계산량을 줄이기 위하여 LP 다항식의 근을 구하지 않고 식 (3)처럼 전체 LP 계수에 γ^i 값을 곱함으로써 결과적으로 시스템의 극점들이 모두 일정 비율로 원점 근처로 이동하도록 하는 방법을 제안하였다[9].

$$H(z/\gamma)=\frac{1}{A(z/\gamma)}=\frac{1}{1-\sum_{i=1}^{P}a_i\gamma^i z^{-i}} \tag{3}$$

여기서 $(0<\gamma<1)$이다.

이 경우 pole filtering이 적용된 LP 켑스트럼은 식 (4)와 같이 쓸 수 있다.

$$pole\ filtered\ Lpcc(n)=\frac{1}{n}\sum_{i=1}^{P}(\gamma p_i)^n=\frac{\gamma^n}{n}\sum_{i=1}^{P}p_i^n=\gamma^n\cdot Lpcc(n) \tag{4}$$

여기서 p_i는 LP 다항식의 i 번째 근이고, P는 다항식의 차수이다.

식 (4)를 살펴보면 결과적으로 켑스트럼 계수에 γ의 지수 승을 곱한 형태가 되는데, 이는 quefrency영역에서의 고주파 영역으로 갈수록 작은 값이 곱해지게 되므로 결국 log spectrum영역에서 빨리 변화하는 성분들을 제거하는 lowpass filter로 해석할 수 있다. 따라서 pole-filtered cepstrum을 구하는 과정은 다음과 같다.

1. 매 프레임 t에 대해서 다음의 과정을 수행한다.

   $pole\ filtered\ Lpcc(t,n)=\gamma^n\cdot Lpcc(t,n)$

   여기서 $Lpcc(t,n)$은 t번째 프레임의 n차 켑스트럼 계수이다.

2. 음성 전체 프레임에 대해 채널 성분을 다음과 같이 추정한다.

   $$Lpcc_h(n)=\frac{1}{T}\sum_{t=1}^{T}pole\ filtered\ Lpcc(t,n)$$

3. 채널 보상된 켑스트럼을 다음과 같이 구한다.

   $$Lpcc(t,n)_{comp}=Lpcc(t,n)-Lpcc_h(n)$$

위 과정에서 $\gamma(0<\gamma<1)$는 실험에 의해 결정되는 상수 값이다.

4. 실험 및 결과

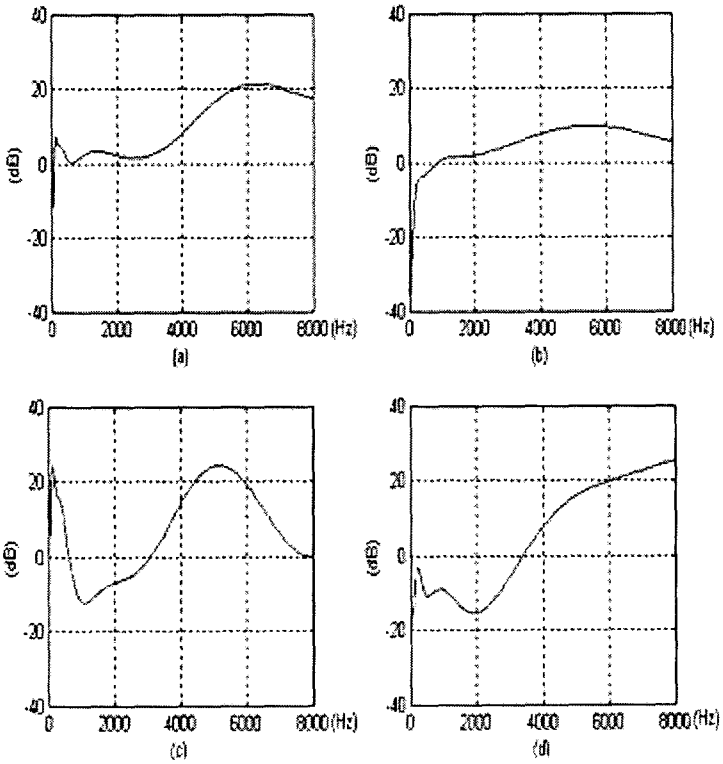
4.1. 음성 데이터베이스

후두암 감별 실험을 위해 한국 장애음성 데이터베이스[10] 및 추가적으로 구축된 음성 데이터를 사용하였으며, 정상인 음성 50개, 후두암 환자 음성 105개, 양성 환자 50개를 사용하였다. 음성 데이터는 모두 남성발성으로 구성되었고 단모음 /아/를 발성하였다. 이들 데이터 중 2/3의 화자는 학습에, 나머지 1/3은 감별 시험에 사용되도록 임의로 5세트를 선정하여 실험하고, 이들의 평균 감별 결과를 사용

하였다. 그리고 기존 장애 음성분석 방법으로 널리 사용되는 Multi-Dimensional Voice Program (MDVP) 분석이 안 되는 후두암 환자 음성 25개도 포함하여 훈련 및 테스트에 사용하였다. 모든 음성은 16kHz로 샘플링 되었고 16bit로 양자화 되었다. 켑스트럼 차수는 12차-20차로 사용하였고 특징 파라미터는 각각 20msec, 30msec, 40msec 크기의 프레임을 10msec씩 이동시키면서 추출하였다.

4.2. 채널 시뮬레이터

후두암 감별용 데이터베이스에 채널에 의한 음성 왜곡을 표현하기 위하여 본 논문에서는 상용 다이내믹 마이크 4가지 종류에 대한 주파수 응답을 FIR 필터 형태로 모델링하여 사용하였다. <그림 4>는 본 논문에서 사용한 4가지 FIR 필터의 주파수 응답이다.



<그림 4> 실험에 사용한 네 가지 마이크 특성 필터의 주파수 응답

4.3. 성능 평가 방법

성능평가를 위해 예측도(predictability)를 구하였다.

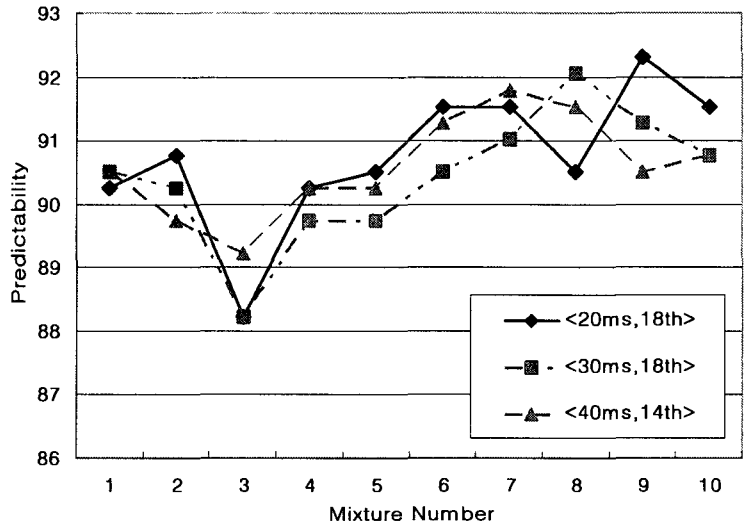
$$\Pr edictability = \frac{True\ Positive + True\ Negative}{Total} \quad (5)$$

여기서 True Positive는 악성을 악성으로 판단한 경우이고, True Negative는 정상 또는 양성을 올바르게 판단한 경우이다.

4.4. 실험결과

Baseline 후두암 감별 실험을 위해서 window 크기를 20msec, 30msec, 그리고 40msec로 변화시키고, 각 window 크기별로 12차에서 20차까지의 LPCC를 추출하면서 실험을 하였다. 실험 결과 20msec 윈도우를 사용한 경우 18차, 30msec를 사용한 경우 18차, 그리고 40msec를 사용한 경우에는 14차가 가장 좋은 성능을 나타내었다. <그림 5>에 window 크기별, 그리고 켑스트럼 차수별로 가장 좋은 성능을 나타내는 경우에 대해 mixture 수에 따른 예측도 성능을 나타내었다.

전체적으로 볼 때 20msec 윈도우, 18차, 그리고 mixture 수가 9인 경우 92.31%로 가장 좋은 성능을 나타내었다. 이후 마이크 채널보상 실험의 경우 위 결과를 baseline으로 사용하여 실험하였다.



<그림 5> 윈도우 크기 및 켑스트럼 차수에 따른 후두암 감별성능 (예측도(%))

<표 1>에 네 가지 마이크 특성에 의해 채널왜곡된 음성에 대해 채널보상 방법에 따른 후두암 감별성능을 나타내었다. 실험에서 α와 γ값은 실험을 통해 최적화하였으며, 최종적으로 각각 0.74와 0.84를 사용하였다. 표에서 "Baseline"의 경우 훈련시 채널보상 방법을 적용하지 않고 실험한 경우이고, "CMS"는 기존의 CMS 방법, "Conv. PFCMS" 및 "Fast PFCMS"는 각각 3.2절에서 method 1과 method 2를 나타낸다.

<표 1> 채널보상 방법에 따른 후두암 감별성능 (예측도(%))

	Baseline	CMS	Conv. PFCMS	Fast PFCMS
Mic1	88.97	80.77	85.64	88.97
Mic2	88.97	80.51	87.95	87.95
Mic3	81.03	78.97	89.23	84.10
Mic4	86.67	79.49	89.23	87.95
Average	86.41	79.94	88.01	87.24

실험결과 일반적인 CMS를 적용한 경우 baseline 성능보다 오히려 떨어지는 것을 볼 수 있는데, 이것은 테스트 데이터가 단모음 /아/로 이루어져 있어서 채널보상 시 채널 정보보다 음성 정보가 많이 제거되기 때문이다. 전체적으로 CMS보다 PFCMS가 성능이 좋고, Fast PFCMS의 경우 baseline보다 평균적으로 우수한 성능을 보이며, Conv. FPCMS의 경우 가장 우수한 성능을 나타내었다. 특히 마이크 3번과 4번의 경우 각각 81.03%에서 89.23%로 43%, 86.67%에서 89.23%로 19%의 인식 오류 감소를 얻었다. 그러나 마이크 1과 2의 경우, PFCMS가 baseline에 비해서는 성능이 동등하거나 오히려 떨어지는 현상을 보였다. 이는 CMS보다는 상대적으로 덜하지만 PFCMS 방법도 채널성분 추정과정에서의 오차로 인해 음성의 왜곡을 초래하며, 그림 4에서 보는 것처럼 마이크의 주파수 특성이 상대적으로 평탄할 경우 PFCMS에 의한 왜곡이 개선 효과보다 클 수도 있기 때문이다. 추후 마이크에 의한 주파수 왜곡이 별로 없는 경우(주파수 특성이 비교적 평탄한 경우) PFCMS의 성능을 더 높이기 위한 방안에 대해 추가적인 연구가 필요하다고 판단된다.

5. 결 론

본 논문에서는 켑스트럼 계수를 기반으로 후두질환 음성의 자동 감별 실험 및 마이크 특성차이에 따른 채널보상에 대한 방법을 검토하였다. Baseline 후두암 감별 성능평가로 window 크기, 켑스트럼 차수, 그리고 mixture 수에 따른 실험결과 20msec 분석윈도우, 18차 켑스트럼, 그리고 mixture 수를 9개 사용한 경우 92.31%로 성능이 가장 우수하였다. 그리고 마이크 특성 차이에 따른 채널 보상 실험을 한 경우 PFCMS를 사용하였을 때 기존의 CMS보다 전체적으로 우수한 성능을 보이고 baseline 결과에 비해서도 평균적으로는 우수한 성능을 보였다. PFCMS를 사용한 결과 baseline에 비해 평균적으로 12%, 기존의 CMS에 비해 40%의 인식 오류율 감소를 얻었다.

참 고 문 헌

[1] P. Lieberman, "Perturbations in vocal pitch", *J. Acoust. Soc. Am.*, 33: pp.597-603, 1961.

[2] S. Iwata, "Periodicities of pitch perturbation in normal and pathologic larynxes", *Laryngoscope,* 82: pp.87-96, 1972.

[3] E. Yumoto, W. J. Gould, T. Baer, "Harmonic-to-noise ratio as an index of the degree of hoarseness", *J. Acoust. Soc. Am.,* 71: pp.1544-1550, 1982.

[4] H. Kasuya, S. Ogawa, K. Mashima, S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice", *J. Acoust. Soc. Am.,* 80(5), Nov., 1986.

[5] Y. Koike. H. Takhashi, T. C. Calcatera, "Acoustic measurements for detecting laryngeal pathology", *Acta Otolaryngol,* 85: pp.105-107, 1977.

[6] Y. Horri, "Jitter and Shimmer in sustained vocal fry phonation", *Folia Phoniatrica,* vol. 37, pp.81-86, 1985.

[7] 김수미, 김형순, 왕수건, 조철우, 양병곤, "GMM 인식기를 이용한 후두질환 음성의 자동 식별 성능 비교", *한국음향학회 학술대회 논문집*, pp.359-362, 2003, 8월.

[8] D. Naik, "Pole filtered cepstral mean subtraction", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp.157-160, May 1995.

[9] Ravi P. Ramachandran and Kevin R. Farrell, "Fast pole filtering for speaker recognition", *IEEE International Symposium on Circuit and System*, vol. 2, pp.365-368 May 2000.

[10] Korean Disordered Speech Database, 창원대학교, 1999.

접수일자 : 2004년 4월 25일
게재결정 : 2004년 6월 7일

▶ 김영국 (Young Kuk Kim)
주소 : 609-735 부산시 금정구 장전동 산30번지 부산대학교 전자공학과
소속 : 부산대학교 전자공학과 음성통신연구실
전화 : 051) 510-1704
FAX : 051) 515-5190
E-mail : ykukim@pusan.ac.kr

▶ 김수미 (Soo Mi Kim)
주소 : 609-735 부산시 금정구 장전동 산30번지 부산대학교 전자공학과
소속 : 부산대학교 전자공학과 음성통신연구실
전화 : 051) 510-1704
FAX : 051) 515-5190
E-mail : noise2@pusan.ac.kr

▶ 김형순 (Hyung Soon Kim)
주소 : 609-735 부산시 금정구 장전동 산30번지 부산대학교 전자공학과
소속 : 부산대학교 전자공학과 음성통신연구실
전화 : 051) 510-2452
FAX : 051) 515-5190
E-mail : kimhs@pusan.ac.kr

▶ 왕수건 (Soo-Geun Wang)
주소 : 602-739 부산시 서구 아미동 1-10 부산대학교 의과대학 이비인후과
소속 : 부산대학교 의대 이비인후과
전화 : 051) 240-7331
E-mail : wangsg@pusan.ac.kr

▶ 조철우 (Cheol-Woo Jo)
주소 : 641-773 경남 창원시 사림동 9 창원대학교 제어계측공학과
소속 : 창원대학교 제어계측공학과
전화 : 055) 279-7552
E-mail : cwjo@sarim.changwon.ac.kr

▶ 양병곤 (Byung-Gon Yang)
주소 : 614-714 부산시 부산진구 가야동 24 동의대학교 영어학과
소속 : 동의대학교 영어학과
전화 : 051) 890-1227
FAX : 051) 890-1222
E-mail : cwjo@sarim.changwon.ac.kr