

켄스트럼 파라미터를 이용한 후두암 검진

부산대학교 의과대학 이비인후과학교실,¹ 의공학교실,² 부산대학교 전자공학과,³
동의대학교 영어영문학과,⁴ 창원대학교 제어계측학과⁵

이원범¹ · 왕수건¹ · 전경명¹ · 권순복¹ · 전계록² · 김수미³ · 김형순³ · 양병곤⁴ · 조철우⁵

= Abstract =

Laryngeal Cancer Screening using Cepstral Parameters

Won-Bum Lee, MD¹, Soo-Geun Wang, MD¹, Kyong-Myong Chon, MD¹,
Soon-Bok Kwon, MS¹, Kye-Rok Jeon, PhD², Soo-Mi Kim, MS³,
Hyung Soon Kim, PhD³, Byong-Gon Yang, PhD⁴, Cheol-Woo Jo, PhD⁵

¹Department of Otolaryngology, College of Medicine, Pusan National University and ²Biomedical Engineering, Busan; and ³Department of Electronics Engineering, Pusan National University, Busan; and ⁴Department of English, College of Humanities, Dong-Eui University, Busan; and ⁵Department of Control & Instrumentation Engineering, Changwon National University Changwon, Korea

Background and Objectives : Laryngeal cancer discrimination using voice signals is a non-invasive method that can carry out the examination rapidly and simply without giving discomfort to the patients. If appropriate analysis parameters and classifiers are developed, this method can be used effectively in various applications including telemedicine. This study examines voice analysis parameters used for laryngeal disease discrimination to help discriminate laryngeal diseases by voice signal analysis. The study also estimates the laryngeal cancer discrimination activity of the Gaussian mixture model (GMM) classifier based on the statistical modelling of voice analysis parameters.

Materials and Methods : The Multi-dimensional voice program (MDVP) parameters, which have been widely used for the analysis of laryngeal cancer voice, sometimes fail to analyze the voice of a laryngeal cancer patient whose cycle is seriously damaged. Accordingly, it is necessary to develop a new method that enables an analysis of high reliability for the voice signals that cannot be analyzed by the MDVP.

To conduct the experiments of laryngeal cancer discrimination, the authors used three types of voices collected at the Department of Otorhinolaryngology, Pusan National University Hospital. 50 normal males voice data, 50 voices of males with benign laryngeal diseases and 105 voices of males laryngeal cancer. In addition, the experiment also included 11 voices data of males with laryngeal cancer that cannot be analyzed by the MDVP. Only monosyllabic vowel /a/ was used as voice data. Since there were only 11 voices of laryngeal cancer patients that cannot be analyzed by the MDVP, those voices were used only for discrimination. This study examined the linear predictive cepstral coefficients (LPCC) and the mel-frequency cepstral coefficients (MFCC) that are the two major cepstrum analysis methods in the area of acoustic recognition.

Results : The results showed that this mel frequency scaling process was effective in acoustic recognition but not useful for laryngeal cancer discrimination. Accordingly, the linear frequency cepstral coefficients (LFCC) that excluded the mel frequency scaling from the MFCC was introduced. The LFCC showed more excellent discrimination activity rather than the MFCC in predictability of laryngeal cancer.

Conclusion : In conclusion, the parameters applied in this study could discriminate accurately even the terminal laryngeal cancer whose periodicity is disturbed. Also it is thought that future studies on various classification algorithms and parameters representing pathophysiology of vocal cords will make it possible to discriminate benign laryngeal diseases as well, in addition to laryngeal cancer.

KEY WORDS : Laryngeal cancer · Voice analysis · Screening test · Cepstral parameters.

논문접수일 : 2003년 10월 28일

심사완료일 : 2003년 11월 1일

책임저자 : 왕수건, 부산광역시 서구 아미동 1-10 부산대학교 의과대학 이비인후과학교실

전화 : (051) 240-7331 · 전송 : (051) 246-8668 E-mail : wangsg@pusan.ac.kr

서 론

음성을 이용한 후두질환의 감별은 고통 없이 검사를 신속 간편하게 할 수 있는 비침습적인 방법으로 1961년 Lieberman이 피치(pitch) 주기 파라미터를 개발하여 후두질환의 감별을 시도한 이래,¹⁾ 신호처리 기술의 발달과 더불어 음성신호의 분석에 의한 후두질환 감별하기 위한 다양한 연구가 진행되어 왔다.²⁻⁷⁾ 그러나 후두암의 병기가 말기에 가까울수록 음성에 잡음 성분이 많이 내재되어 음성분석에 의한 감별이 많은 오류를 유발하기 때문에 아직도 만족할 만한 방법이 없는 실정이다.

후두질환 감별용 음성 분석방법 중 대표적인 것으로 Multi-dimensional voice program(MDVP) 분석에 의한 jitter나 shimmer 파라미터 등을 들 수 있으나,⁸⁾ 이들의 경우 주기성이 크게 훼손된 후두암 말기의 음성 데이터에 대해서는 분석 자체가 불가능해지는 문제점이 있다. 본 논문에서는 MDVP 분석이 불가능할 정도로 주기성이 훼손된 음성에 대해 효과적인 감별이 가능한 신뢰도 높은 분석방법으로서, 음성인식 분야에서 널리 사용되어 온 켈스트럼(cepstrum) 파라미터 분석방법을 검토하였다. 실제로 음성인식 분야에 주로 사용되는 특징 파라미터는 linear predictive cepstral coefficients(LPCC)와 mel-frequency cepstral coefficients(MFCC)이며,⁹⁾¹⁰⁾ 본 논문에서는 이들과 더불어 MFCC의 변형된 형태인 linear frequency cepstral coefficients(LFCC)를 함께 검토하였다. 이들 파라미터 분석 결과로부터 정상 음성 및 양성 후두질환 음성, 그리고 악성 후두질환 음성을 Gaussian mixture model(GMM)로 모델링한 GMM분류기¹⁰⁾를 통해 후두암 여부를 감별하고자 본 연구를 시행하였다.

연구대상 및 방법

1. 연구 대상

본 논문에서는 후두암 감별 실험을 위해 부산대학교병원 이비인후과에서 수집한 정상 남자 음성 데이터 50개, 양성 후두질환 남자 음성 데이터 50개 및 남성 후두암환자 음성 데이터 105개를 사용하였다. 그 외에도 기존의 MDVP (Multi-dimensional voice program) 분석이 불가능한 후두암 환자 남성 음성 데이터 11개도 포함하였다. 음성 데이터는 단모음 /아/ 발성만을 사용하였고, 정상인과 양성 후두질환 환자, 그리고 MDVP 분석이 가능한 후두암 환자 음성 데이터 중 2/3는 학습에, 나머지 1/3은 감별실험에 사

용하였다. MDVP로 분석되지 않는 후두암 환자의 음성 데이터는 학습에는 포함시키지 않고, 모두 감별실험에만 사용하였다.

2. 실험 방법

실험결과의 일관성을 높이기 위해 학습 및 감별에 사용한 음성 데이터는 다수의 세트를 무작위로 선정하여 5번 실험을 수행한 후 이들의 평균 감별결과를 계산하였다. 음성 데이터의 샘플링 주파수는 16kHz이며, 양자화 비트 수는 16비트를 사용하였다. LPCC, MFCC 및 LFCC 등 켈스트럼 파라미터를 얻기 위해서 프레임에 Hamming 윈도우를 사용하였으며, pre-emphasis 계수는 0.97로 하였다.

GMM 분류기의 훈련은 프레임 단위로 수행하고, 감별 과정 역시 1차적으로 프레임 단위로 확률을 계산한 다음 전체 구간에 대해 누적 log 확률값을 구해서 발화 단위로 감별 판정을 하였다. 실험 결과는 후두암 전체 음성 데이터에 대해, 그리고 MDVP 분석이 불가능한 후두암 음성 데이터에 대해 민감도(sensitivity), 특이도(specificity) 및 예측도(predictability)를 구하였다.

이하에 본 논문에서 사용한 켈스트럼 파라미터인 LPCC, MFCC 및 LFCC에 대해 간단히 설명하고, 이와 더불어 GMM 분류기에 대해서도 언급한다.

3. Linear Prediction Cepstral Coefficients(LPCC)

선형예측(linear prediction) 분석 방법은 음성신호처리에 널리 사용되는 분석 방법으로서, 이 방법은 현재의 음성 샘플이 과거 p개의 음성 샘플의 선형 조합에 의해 예측할 수 있다는 데에 기반을 두고 있다.⁹⁾

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p) \quad (1)$$

여기서 분석하고자 하는 특정한 음성 프레임 내에서 예측오차 에너지를 최소화하는 계수들을 선형예측 계수 또는 LPC 계수라고 하며, 이들 계수에 의해 음성발생기관의 전달함수는 다음 식과 같이 구해질 수 있다.

$$H(z) = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (2)$$

따라서, 선형예측 계수는 음성발생기관의 모델에 기반하여 음성 스펙트럼을 표현하는 수단으로 사용될 수 있으며, 이를 음성의 로그 스펙트럼을 표현하는 계수로 변환한 것을 선형예측 켈스트럼 계수 또는 LPC 켈스트럼 계수(LPCC)라고 한다. 선형예측 계수로부터 LPC 켈스트럼 계수로 변환하는 과정을 수식으로 표현하면 다음과 같다.

$$C_m = a_m + \sum_{n=1}^{k-1} \frac{k}{m} c_k a_{m-k}, \quad 1 \leq m \leq p \quad (3)$$

4. Mel-Frequency Cepstral Coefficients(MFCC)

음향심리학의 연구에서 물리적 주파수와 인간이 인지하는 주관적인 주파수인 피치(pitch) 사이의 관계는 선형적이 아님이 밝혀졌으며, 이에 따라 인간이 느끼는 주관적인 주파수의 단위인 mel이 정의되었다.¹⁰⁾ 음성인식을 위한 특징 파라미터를 추출할 때 사람이 인지하는 주관적 주파수의 특성을 반영한 mel 주파수 척도에 따라, 대역필터군(bandpass filterbank)를 비선형으로 구성하여 사용하면 좋은 성능을 얻을 수 있다. 이러한 대역필터군을 통해 분석된 음성 스펙트럼으로부터 구한 켈스트럼 계수를 MFCC라고 한다.¹¹⁾ MFCC 분석에 사용되는 mel 주파수 척도를 적용하여 구성된 삼각형 모양의 대역필터군의 예는 아래와 같다(Fig. 1).

MFCC를 구하는 과정을 정리하면 다음과 같다. 음성이 들어오면 먼저 음성의 고주파 부분의 감쇄를 보상해주는 pre-emphasis 과정을 거치고, 그 다음 고속 Fourier 변환(FFT)을 통해 전력 스펙트럼을 구한다. 여기에 삼각 대역필터군이 적용된 다음 로그를 취하고, 마지막으로 다음 식과 같은 이산 코사인 변환(discrete cosine transform, DCT)을 거쳐 MFCC가 구해진다(Fig. 1).

$$C_m = \sqrt{\frac{2}{N}} \sum_{i=1}^N \left\{ \log(X[i]) \cos\left(\frac{2\pi m}{N} \left(i - \frac{1}{2}\right)\right) \right\} \quad (4)$$

여기서 $X[i]$ 는 i -번째 mel 주파수 대역필터의 출력이고, N 은 필터군의 개수이며, m 은 켈스트럼의 차수를 의미한다.

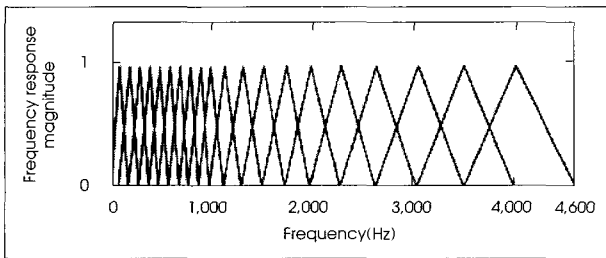
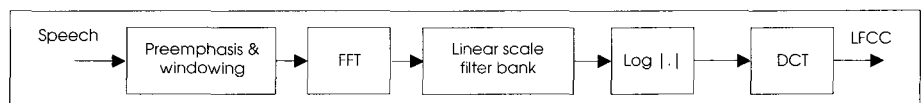


Fig. 1. Mel-frequency scaled bandpass filterbank for extracting MFCC.

Fig. 2. Extraction of MFCC.



Fig. 3. Extraction of LFCC.



아래의 그림은 MFCC 추출과정의 구성도이다(Fig. 2).

5. Linear Frequency Cepstral Coefficients(LFCC)

LFCC는 MFCC와 마찬가지로 대역필터군 분석에 의해 구해지는 켈스트럼 계수이다. LFCC가 MFCC와 유일하게 다른 점은 Fig. 1에서와 같이 삼각 필터군을 구성할 때, mel scale을 사용하지 않고 선형 scale을 사용한다는 것이다. 즉, LFCC에서는 MFCC의 경우와 달리 삼각 필터들의 중심주파수가 주파수 영역에서 동일한 간격으로 배치된다.

음성인식 연구에 따르면 mel scale을 사용하는 MFCC가 청각기관의 특성을 보다 잘 반영하기 때문에 LFCC에 비해 성능이 우수하다고 보고되었다.¹¹⁾ 그러나 mel scale이 음소들의 변별력을 높여서 음성인식 성능향상에는 기여하더라도, 후두암 여부의 변별력 향상에도 유용하다는 근거는 없으며, 이에 관한 기존 연구도 없는 실정이다. 따라서 본 논문에서는 후두암 음성감별에 mel scale의 유용성 여부를 파악하기 위해 MFCC와 LFCC의 성능을 실험을 통해 비교하였다. LFCC 특징 파라미터를 추출하는 과정은 다음과 같다(Fig. 3).

6. Gaussian Mixture Model(GMM) 분류기를 이용한 후두암 선별 시스템

켈스트럼 파라미터들을 이용한 후두암 음성감별을 위해 본 논문에서는 Gaussian mixture model(GMM) 분류기를 사용하였다. GMM 분류기는 음성감별 실험에 널리 사용되는 다층 퍼셉트론 신경회로망 분류기에 비해 사용 및 해석 면에서 유리하며, 선행연구를 통해 성능면에서 대등한 것으로 나타났다.¹²⁾

GMM은 임의의 확률분포를 식 (5)와 같이 M 다변수의 가중 Gaussian mixture 분포로 표현하며, 이를 그림으로도 시하면 다음과 같다(Fig. 4).¹⁰⁾

$$p(x | \lambda) = \sum_{i=1}^M a_i b_i(\vec{x}) \quad (5)$$

여기서

$$b_i(\vec{x}) = N(\vec{x}, \mu_i, \Sigma_i) \quad (6)$$

이고, \vec{x} 는 특징 벡터이며, μ 와 Σ 는 특징 벡터의 평균과

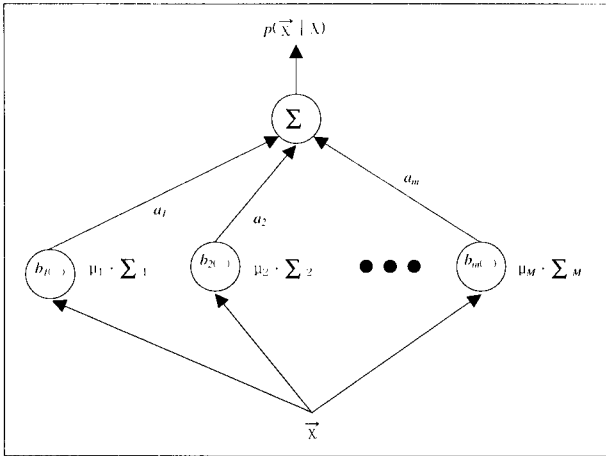


Fig. 4. GMM with M mixtures.

공분산(covariance) 행렬, 그리고 λ 는 GMM의 파라미터 모델이다. 그리고, 정규분포 $N(\cdot)$ 은

$$N(\vec{x}, \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_i|^{1/2}}} \exp[-\frac{1}{2}(\vec{x} - \mu_i)^T \Sigma_i^{-1} (\vec{x} - \mu_i)] \quad (7)$$

와 같다. 식 (5)에서의 a_i 는 전체 벡터의 분포에서 특정 i -번째 Gaussian 분포가 차지하는 가중치이고, 식 (7)에서의 D 는 특징 벡터의 차원이다.

주어진 특징 벡터 \vec{x} 에 대해서 GMM의 파라미터들은 Expectation-Maximization (EM) 알고리즘을 통해서 추정 가능하다고 한다.¹⁰⁾

결 과

3가지 켈스트럼 파라미터(LPCC, MFCC 및 LFCC)에 의한 후두암 감별실험 결과가 Fig. 5에 나타나 있다. 그림에서 볼 수 있듯이 GMM 분류기에서의 mixture 수를 1에서 10까지 증가시켜 가면서 실험을 하였다(Mixture의 수가 늘어나면 모델의 정밀성이 높아지지만, 제한된 데이터에 대해 추정해야 할 모델 파라미터의 수가 증가되므로 추정된 파라미터의 신뢰도가 떨어질 수 있다). 실험 결과, LPCC, MFCC 및 LFCC를 이용한 최고 예측도가 각각 91.3%, 88.7% 및 92.1%로 나타났으며, LFCC의 성능이 가장 우수함을 알 수 있었다. 민감도와 특이도에 대한 평가도 함께 수행하였으며, 민감도 면에서는 MFCC가, 그리고 특이도 면에서는 LFCC와 LPCC가 상대적으로 우수한 성능을 나타내었다.

감별실험용 후두암 음성으로 MDVP 분석이 불가능한 데

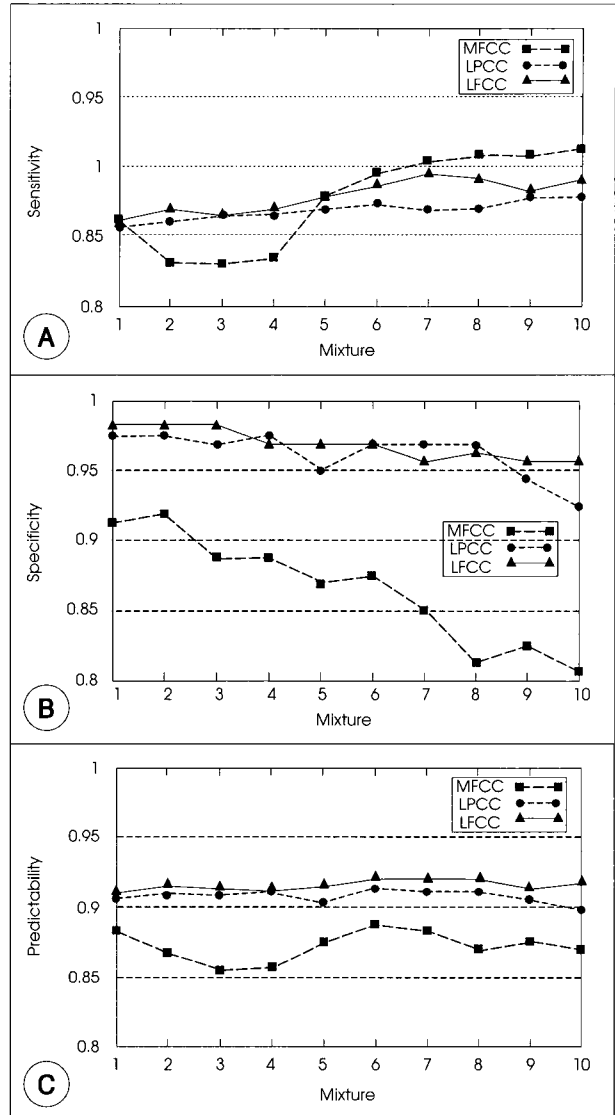


Fig. 5. Performance comparison of laryngeal cancer screening according to various cepstral parameters. A : Sensitivity, B : Specificity, C : Predictability.

이터만을 이용한 경우에도 LPCC, MFCC 및 LFCC를 이용하여 각각 90.2%, 90.2% 및 90.7%의 예측도를 얻었으며, MDVP 분석불능 데이터를 모델 훈련에는 사용하지 않고 감별실험에만 사용했음을 고려할 때 상당히 높은 성능이 얻어졌다고 판단된다.

고 찰

음성분석법은 비침습적이며, 객관적인 검사법으로 조기 성대 병변 뿐 만 아니라 수술, 약물치료, 음성재활 치료 전후의 효과판정에 유용하다. 음성분석에 널리 사용되는 방법은 Kay사의 CSL(Computerized Speech Lab, 4,300B) 기

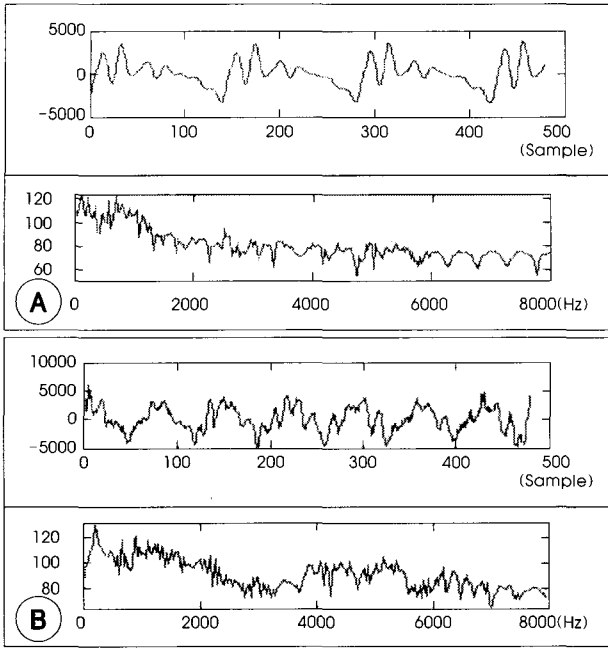


Fig. 6. Examples of speech waveforms(upper) and spectra(lower) of normal voice and voice with laryngeal cancer(vowel /a/). A : Normal voice, B : Voice with laryngeal cancer.

중의 음성분석 소프트웨어인 MDVP가 대표적이며 그 외 CM Instrument사의 Sound Scope, Tiger Electronics사의 Dr. Speech, Praat 등을 이용한 경우에도 거의 비슷한 파라미터의 분석이 가능하다. MDVP는 기본주파수 변화, 진폭의 변화, 떨림(tremor) 여부, 잡음성분의 비율 등 4가지 범주의 33개 파라미터들로 구성되어 있으며, 이 중 일부 파라미터들이 기질적 병변에 의한 후두질환 감별에 유용하다.¹³⁾

기질적 음성장애 초기 병변의 특징은 잡음 성분이 약간 증가되고, 고음역 배음(high frequency harmonics)의 크기가 감소하지만 피치의 주기성은 거의 정상(normal periodicity)을 유지한다. 그러나 병변이 진행하면 기본주파수의 주기와 주기 내의 음성 크기(amplitude)의 동요(perturbation)가 생기며, 가끔 기본주파수와 발성 크기의 감소로 음역(voice range)이 줄어들게 된다. 그외 제 1 배음이 뚜렷해지고(dominating 1st harmonic), 1kHz 이상의 고음역 배음은 거의 소실되며, 배음과 배음사이에 잡음성분이 출현하게 된다(subharmonic component).¹⁴⁾ 이런 변화는 후두암 환자의 음성에서 더 뚜렷한 변화를 나타낸다. 후두암 환자의 음성은 잡음 성분이 많으므로 스펙트럼의 고주파 성분이 상대적으로 높고, 주기성이 심하게 떨어지므로 주파수 영역에서의 배음 성분들이 거의 드러나지 않는다. 그러므로 음성 스펙트럼의 포락선 정보만으로도 정상인과 후두암 환자의 구별이 가능함을 알 수 있다(Fig. 6).

음성신호의 분석은 전술한 특정 후두질환을 잘 반영할 수

있는 분석 파라미터가 일차적으로 중요하지만 이들의 분석 결과를 이용하여 후두질환을 감별하는 분류기(classifier)에 대한 연구가 동반되어야 가능하다. 이를 위하여 여러 분류기법들이 적용되었는데, 지금까지 후두암 감별 연구에 사용되어 온 분류기로는 다층 퍼셉트론 신경회로망 분류기¹⁵⁻¹⁷⁾와 다층 퍼셉트론 신경회로망과 대등한 성능을 보이면서도 사용 면에서 편리한 GMM 분류기,¹²⁾ 그외 k-nearest neighbor(k-NN) 분류기,¹⁴⁾ Hidden Markov model(HMM) 분류기 등이 있다.¹⁸⁾

GMM 분류기는 복수의 Gaussian 분포의 가중합의 형태로 분류패턴을 모델링하는 것에 기반을 둔 것으로서, Gaussian 분류기¹⁶⁾의 확장된 형태의 모수적 분류기이다. 모수적 분류기는 분류패턴의 특성을 잘 표현할 수 있는 모델을 설정한다는 전제 하에서 우수한 성능을 나타낸다.¹⁶⁾ Gaussian 분류기가 unimodal한 분포에 기반을 두고 있어서, 선형적으로 분리가능한(linearly separable)한 패턴분류 문제에만 적용 가능한 데 반해서, GMM 분류기는 mixture 수, 즉 Gaussian 분포의 개수를 적절히 조절하면 다양한 종류의 분포를 표현할 수 있는 장점이 있어서, 음성인식과 화자인식을 비롯한 여러 분야에 널리 활용되고 있다. 그러나 GMM 분류기도 최적의 mixture 수를 실험적으로 결정해야 한다는 점과 신경회로망과 마찬가지로 모델 파라미터 추정과정에서 sub-optimal하다는 문제점이 있다.

주기성 분석 파라미터를 이용한 후두암 감별을 보면 Kasuya 등은 잡음분석 파라미터를 이용하여 성문암 제 1기의 경우 77.4%, 제 2~4기 100%의 결과를 보고하였고,⁵⁾ Wang 등은 MDVP 파라미터의 결과를 정준분석방법을 이용하여 감별율이 87%였다고 보고하였다.¹⁹⁾ 그러나 MDVP의 경우와 같이 음성신호의 주기판별과 개별주기 단위로의 파형 구분을 전제로 추출되는 파라미터들은 말기 후두암 환자의 음성과 같이 음성신호의 주기성이 크게 훼손되어 있으면 파라미터의 분석 자체가 불가능한 경우가 종종 발생한다. 따라서 MDVP 분석 방법은 모든 병기의 후두암 음성에 적용할 수 있는 적절한 분석방법이 되지 못하며, MDVP 분석 이외의 다른 대안이 필요하다. 최근 음성인식 분야에서 유용하게 사용되는 캡스트럼 분석방법을 후두질환의 감별에 이용하는 연구가 진행되고 있다.

캡스트럼이란 음성 스펙트럼의 포락선 정보를 표현하기 위한 하나의 방법으로 로그 스펙트럼의 역 Fourier 변환을 말하며,⁹⁾ 일반적으로 캡스트럼 분석은 convolution에 의해 결합된 두 신호를 분리하는 데에 효과적인 도구로 알려져 있다. Jo 등은 자체 개발한 캡스트럼 파라미터인 Harmonic-to-Noise Ratio : Residual(HNRR)과 기존의 음향학적 파

라미터를 함께 이용하여 신경회로망 분류기로 분석한 결과 정상과 비정상의 감별 성능 94.5% 및 양성 후두질환과 후두암의 감별성능이 91.7%였다고 보고하였다.²⁰⁾ Martinez와 Rufiner는 주파수 영역에 기초를 둔 여러 가지 음성신호처리 방법 즉 켈스트럼(cepstrum), 멜켈스트럼(mel-cepstrum), 델타 켈스트럼(delta cepstrum) 및 델타 멜켈스트럼(delta mel-cepstrum), FFT 방법을 이용하여, 얻어진 데이터는 인공신경망(artificial neural network)을 통해 정상과 병적 음성의 분류 및 정상과 이중 음성(diphonia), 거친 음성의 분류를 시도한 결과 켈스트럼 방법이 감별율 91.3%로 가장 우수하였다고 보고하였다.¹⁷⁾

본 논문에서는 MDVP 분석이 불가능한 말기 후두암음성 데이터에 대해서도 효과적으로 감별할 수 있는 신뢰도 높은 음성분석 방법으로 LPCC, MFCC 및 LFCC를 비교하였다. 그 결과, MFCC에서 1kHz 이하의 낮은 주파수 스펙트럼의 특성에는 민감하고 1kHz 이상의 주파수 스펙트럼에는 상대적으로 둔감한 특성을 가지도록 mel 주파수 scaling을 적용하는 과정이 후두암의 감별에서는 유용성이 부각되지 못했다. 실험 결과 MFCC에서 mel 주파수 scaling 부분을 제외시킨 LFCC가 MFCC보다 특이도 및 예측도 측면에서 우수하였고, 민감도에서만 MFCC가 LFCC보다 우수하였다.

결론적으로 말기 후두암과 같이 주기성이 훼손된 경우라도 본 연구에서 적용한 파라미터들로서 정확한 감별이 가능하였으나, 각각의 성대병변의 병태생리를 대변할 수 있는 파라미터와 다양한 분류 알고리즘에 대한 연구를 시행하면 후두암 이외에도 양성 후두질환의 감별도 가능할 것으로 사료되었다.

결 론

후두암의 음성분석을 위해 MDVP 분석 파라미터들이 널리 사용되고 있으나, 주기성이 크게 훼손된 심한 후두암 환자의 음성에 대해서는 분석 자체가 불가능한 경우가 종종 있다. 따라서 MDVP 분석이 불가능한 음성 신호에 대해서도 신뢰성이 높고 분석이 가능한 음성분석 방법이 필요하며, 본 논문에서는 이 문제의 해결을 위해 켈스트럼 파라미터들을 사용하였다.

후두암 감별 실험 결과, LPCC, MFCC 및 LFCC의 세 가지 켈스트럼 파라미터 중 예측도 관점에서 LFCC의 성능이 가장 우수하였다. 민감도와 특이도에 대한 평가도 함께 수행하였으며, 민감도 면에서는 MFCC가, 그리고 특이도 면에서는 LFCC와 LPCC가 상대적으로 우수한 성능을 나타내었다. 감별실험용 후두암 음성으로 MDVP 분석이 불가

능한 데이터만을 이용한 경우에도 LPCC, MFCC 및 LFCC 모두 최고 90% 대의 예측도를 얻었으며, MDVP 분석불능 데이터를 모델 훈련에는 사용하지 않고 감별실험에만 사용했음을 고려할 때 상당히 높은 성능이 얻어졌다고 판단된다.

결론적으로 말기 후두암과 같이 주기성이 훼손된 경우라도 본 논문에서 적용한 켈스트럼 파라미터를 통해 비교적 정확한 감별이 가능하였다. 앞으로 후두암 감별 성능을 더욱 향상시키기 위하여 후두암 감별에 효과가 있는 다양한 음향 파라미터들을 켈스트럼 파라미터와 함께 적용하는 방안 등에 대해 계속 연구할 필요가 있다고 사료되며, 후두암 이외에도 양성 후두질환의 감별도 함께 적용할 예정이다.

중심 단어 : 켈스트럼 변수 · 후두암 검진.

본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임(02-PJ1-PG10-31401-0005).

REFERENCES

- 1) Lieberman P. *Perturbation in vocal pitch. J Acoust Soc Am* 1961;33: 597-603.
- 2) Iwata S. *Periodicities of pitch perturbation in normal and pathologic larynges. Laryngoscope* 1972;82:87-96.
- 3) Koike Y, Takahashi H, Calcaterra TC. *Acoustic measurements for detecting laryngeal pathology. Acta Otolaryngol* 1977;85:105-17.
- 4) Yumoto E, Gould WJ, Baer T. *Harmonic-to-noise ratio as an index of the degree of hoarseness. J Acoust Soc Am* 1982;71:1544-50.
- 5) Kasuya, Ogawa S, Mashima K, Ebihara S. *Normalized noise energy as an acoustic measure to evaluate pathologic voice. J Acoust Soc Am* 1986;80:1329-34.
- 6) Boyanov B, Hadjitodorov S. *Acoustic analysis of pathological voices. IEEE Engineering in Medicine and Biology*;1997. p.74-82.
- 7) Juan I. Godino L, Santiago AN, Carlos HE, Mercedes FR, Pedro G. *On The Selection of Meaningful Speech Parameters Used by a Pathologic/Nonpathologic Voice Register Classifier. In Proc Eurospeech*;1999. p.563-6.
- 8) *Multi-Dimensional Voice Program (MDVP)-Model 4305 Operations Manual, Kay Elemetrics Corp*;1993.
- 9) Rabiner LR, Juang BH. *Fundamentals of Speech Recognition. Prentice Hall: Englewood Cliffs, New Jersey*;1993. p.97-121.
- 10) Huang X, Acero A, Hon HW. *Spoken Language Processing. Prentice-Hall PTR: Upper Saddle River, New Jersey*;2001. p.275-314.
- 11) Davis SB, Mermelstein P. *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In IEEE Trans Acoust Speech and Sign Process. ASSP-28*;1980. p.357-66.
- 12) Kang HM, Kim SM, Kim YS, Kim HS, Jo CW, Yang BG, et al. *Performance comparison of automatic detection of laryngeal diseases by voice. Malsori* 2003;45:35-46.
- 13) Juan I. Godino L, Santiago AN, Carlos HE, Mercedes FR, Pedro G. *On The Selection of Meaningful Speech Parameters Used by a Pathologic/Nonpathologic Voice Register Classifier. In Proc Eurospeech*;1999. p.563-6.
- 14) Hadjitodorov S, Mitev P. *A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. Med Eng Phys* 2002;24:419-29.
- 15) Jo CW, Wang SG, Yang BG. *A study on the diagnosis of laryngeal diseases by acoustic signal analysis. Korean J Speech Science* 1999;5: 151-65.

- 16) Juan I. Godino L, Santiago AN, Carlos HE, Mercedes FR, Pedro G. *On The Selection of Meaningful Speech Parameters Used by a Pathologic/Nonpathologic Voice Register Classifier. In Proc Eurospeech:1999. p.563-6.*
- 17) Martinez CE, Rufiner HL. *Acoustic analysis of speech for detection of laryngeal pathologies. IEEE Int Conf EMBS:2000. p2369-72.*
- 18) Gavidia-Ceballos L, Hansen JHL. *Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection. IEEE Trans Biomed Eng 1996;43:373-83.*
- 19) Wang SG, Baek MJ, Yang BG, Jo CW, Park HM, Kweon SB. *et al. Acoustic parameters for the early detection and differential diagnosis of pathologic voice. Korean J Otolaryngol 1999;42:1561-7.*
- 20) Jo CW, Kim KI, Kim DH, Kwon SB, Kim KR, Kim YJ. *et al. Classification of pathological voice from ARS using neural network. Korean J Speech Sciences 2001;8:61-71.*