2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

한국어 음성 인식 시스템을 위한 MEL-LPC 분석 방법과 LPC-MEL 분석 방법의 비교

김 주 곤*, 김 범 국**, 정 호 열*, 정 현 열*
*영남대학교 정보통신공학과, **대구과학대학 정보전자통신 계열

Comparison of MEL-LPC and LPC-MEL Analysis Method for the Korean Speech Recognition Systems.

Joo-Gon Kim*, Bum-Guk Kim**, Ho-Youl Jung*, Hyun-Yeol Chung*
*Department of Information & Communication Eng., Yeungnam University
**Informational Electronics & Communication Division, Teagu Science College

요 약

본 논문에서는 한국어 음성인식 시스템의 성능 향상을 위해 청각 주파수 분해능을 가진 MEL-LPC Cepstrum을 음소단위의 HMM(Hidden Markov Model)을 기반으로 하는 인식 시스템에 적용하여 그 결과를 비교 검토하였다.

선형예측(LP) 분석 후에 후처리로서 주파수를 왜곡시킨 LPC-MEL 분석이 계산량이 적고 효과적이라 일반적으로 많이 사용되고 있으나 주파수 분해능은 많이 개선되지 않는다. 따라서 본 논문에서는 주파수 분해능을 개선하기 위해, 원 음성신호로부터 직접적으로 멜 주파수로 왜곡시킨 후 선형 예측 분석을 수행하는 MEL-LPC 분석방법을 이용한 음소기반의 화자 독립 음성인식 시스템을 구성하여 기존의 LPC-MEL 분석방법과 비교실험을 통하여 MEL-LPC 분석방법의 유효성을 검토하였다.

실험에 사용한 음성 데이터베이스는 음소 및 단어 인식실험에서는 ETRI 445단어 DB, 연속 숫자음인식 실험에서는 KLE 4연속 숫자음 DB를 사용하였다. 화자 독립 음소인식 실험의 경우, 묵음을 제외한 47개의 유사 음소에 대하여 4상태 3출력의 Left-to-Right 모델을 이용하였다. 단어 및 연속 숫자음 인식 실험의 경우, 유한상태 네트워크에 의한 OPDP법을 이용하였다.

화자 독립 음소, 단어 및 4연속 숫자음 인식 실험결과, 기존의 LPC-MEL Cepstrum을 사용한 경우보다 MEL-LPC Cepstrum을 사용한 경우가 더 높은 인식률을 나타내어 한국어 음성인식 시스템에서 MEL-LPC 분석방법의 유효성을 확인할 수 있었다.

Ⅰ. 서 론

음성인식 시스템에서 효과적인 음성 특징 추출은 가장 중요한 issue 중에 하나이다. 음성인식 시스템에 사용되는 파라미터로는 MFCC(Mel Frequency Cepstral Coefficients), PLP(Perceptual Linear Predictive), LPC(Linear Predictive Coefficients)등과 이들의 개선된 형태의 파라미터들이 있다. 선형예측 분석 방법은 스펙트럼 포락을 All Pole Model로 표현하여 최적의 파라미터를 비교적 적은 계산량으로 안정하게 추출하기 때문에 많이 사용되었다. 이후 스펙트럼 포락과 피치(Pitch: 기본 주파수)를 분리하여 추출하는 Cepstrum방법이 개발되었고 현재에는 인간의 청각 특성을 고려한 Mel-cepstrum과 동적 특징을 표현한 회귀계수(Regressive Coefficients)가 널리 사용되고 있다. 선형예측(LP) 분석 후에 후처리로서 주파수를 왜곡시

킨 LPC-MEL 분석방법[5,6]이 계산량이 적고 효과적이라 일반적으로 많이 사용되고 있으나 주파수 분해능은 많이 개선되지 않는다.

주파수 분해능을 개선하기 위해, 원 음성신호를 직접적으로 멜 주파수로 왜곡시킨 후 선형 예측 분석을 수행하는 방법[1]이 제안되어졌으나 높은 계산량 때문에 거의 사용되지 않았다. 최근 일반적인 선형예측 분석 방법과 비교해서 단지 두 배의 계산량으로 MEL-LPC cepstrum을 얻는 방법이 제안되었다[2].

따라서 본 논문에서는 한국어 음성인식 시스템의 성능 향상을 위해서 LPC-MEL 분석방법과 MEL-LPC 분석방법을 이용하여 음소모델을 구성하고, 음소인식과 단어 및 연속숫자음 인식 실험을 통하여 두 분석 방법을 비교 검토 하고자 한다.

이를 위하여 음성자료는 한국 전자통신 연구소(ETRI)에서 채록한 445단어(ETRI 445)와 국어공학 연구소(KLE)의 4연속 숫자음 DB를 사용한다. 인식의 기본 단위로는 48개의 유사음소단위(PLUs)를 음소모델로 사용하며, 단어 및 연속숫자음 인식 실험을 위해서는 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP법[3]을 이용한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서 MEL-LPC 분석방법에 대해서, Ⅲ장에서는 인식 실험 및 고찰을 통하여 MEL-LPC 분석방법의 유효성을 확인하고 마지막 Ⅳ장에서 결론을 맺는다.

Ⅱ. MEL-LPC 분석방법

인간의 청각능력은 음의 크기에 대하여 근사적으로 대수적인 특성을 나타내며, 주파수 분해능은 1kHz이하의 낮은 주파수영역에서는 선형적이고 그 이상의 주파수영역에서는 대수적인 멜 척도(Mel-scale) 특성을 가진다. 이러한 특성을 이용하여 음성의 특징 파라미터로서 Mel-cepstrum이 많이 사용되고 있다.

2.1 LPC-MEL Cepstrum 계수 추출

음성의 단구간에 대하여 LPC 분석을 수행한 후 얻어지는 Mel-cepstrum 계수 $\{M_k\}$는 LPC 켑스트럼 계수 $\{c_k\}$에서 근사적으로 식 (1)과 같은 bilinear 변환을 이용한다.

$$\tilde{z}^{-1}=\frac{(z^{-1}-\alpha)}{(1-\alpha z^{-1})} \qquad 0<\alpha<1 \tag{1}$$

여기서 위상특성은 아래와 같다.

$$\Omega=\omega+2\arctan\left[\frac{\alpha\sin(\omega)}{1-\alpha\cos(\omega)}\right] \tag{2}$$

이를 이용하여 LPC-MEL Cepstrum은 아래의 식으로 구해진다.

$$M_k[m]= \tag{3}$$
$$\begin{cases} c[-m]+\alpha\cdot M_k[m-1] & k=0 \\ (1-\alpha^2)\cdot M_{k-1}[m-1]+\alpha\cdot M_k[m-1] & k=1 \\ M_{k-1}[m-1]+\alpha(M_k[m-1]-M_{k-1}[m]) & k>1 \end{cases}$$
$$(m=\cdots,-2,-1,0)$$

여기서 샘플링주파수가 6.67kHz, 8kHz, 10kHz, 16kHz일 경우, α를 각각 0.28, 0.31, 0.35, 0.45로 두면, 쉽게 멜 켑스트럼을 근사적으로 구할 수 있다.

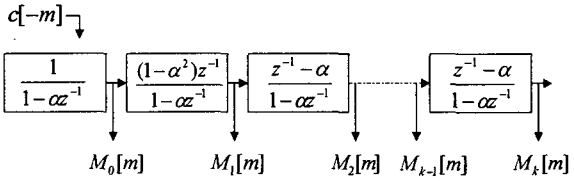

그림 1. LPC-MEL Cepstrum 추출 방법

2.2 MEL-LPC Cepstrum 계수 추출

원 음성신호를 직접적으로 멜 주파수로 왜곡시킨 후 선형 예측 분석을 수행하여 MEL-LPC Cepstrum을 얻기 위한 직선 주파수 축 상에서의 inverse filer는 다음과 같다[2].

$$A_w(z)=\tilde{A}_w(\tilde{z})=\sum_{n=0}^{p}\tilde{a}_{w,n}\tilde{z}^{-n} \tag{4}$$

Durbin's algorithm으로 식 (4)를 계산하기 위하여 다음의 Mel Autocorrelation coefficients를 사용한다.

$$r_w[k]=\sum_{n=0}^{N-1}x[n]y_k[n] \tag{5}$$

여기서 $x[n]$은 원 음성 샘플이고 $y_k[n]$는,

$$y_k[n]=\alpha\cdot(y_k[n-1]-y_{k-1}[n])+y_{k-1}[n-1] \tag{6}$$
$$(n=0,\cdots,N-1,k=1,\cdots,p)$$

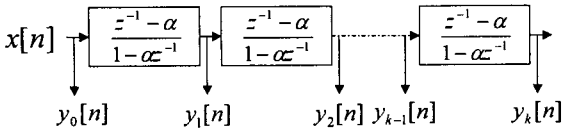

그림 2. MEL-LPC Cepstrum을 얻기 위한 Mel Autocorrelation

2.3 회귀 계수 추출

음성으로부터 특징을 추출할 때 스펙트럼 내의 순시적인 변화를 나타내는 동적 특징 파라미터로서 회귀계수가 사용된다. 회귀계수 $R_m(t)$는 시간 t를 중심으로 $2\delta+1$ 폭 만큼의 단위로 계산하여 구해진다[3].

$$R_m(t) = \frac{\sum_{n=-\delta}^{\delta} nC_m(t+n)}{\sum_{n=-\delta}^{\delta} n^2} \qquad (7)$$

여기서 $C_m(t)$는 t 번째 프레임의 m번째 정적 파라미터의 계수값이고 $R_m(t)$는 여기에 해당하는 회귀계수값을 의미한다.

III 인식 실험 및 고찰

3.1 음성자료 및 분석

화자 독립 음성 인식 실험을 위한 음성자료는 한국전자통신 연구소(ETRI)에서 구축한 한국인 남성 22인의 2회 발성한 445단어(ETRI 445)데이터 중에서 5인이 발성한 단어로부터 추출한 음소로 표준 패턴을 구성하고 학습에 참여하지 않은 3인의 화자가 발성한 단어를 인식 실험에 사용한다. 연속 숫자음 인식 실험을 위해 국어공학연구소(KLE)의 4연속 숫자음 DB 중에서 남성 20인이 발성한 4연속 숫자음을 모델학습에 사용하고, 학습에 참여하지 않은 남성 10인의 화자가 발성한 4연속 숫자음을 평가용 자료로 사용한다. 샘플링 주파수 16kHz인 음성 데이터를 Pre-emphasis 필터를 통과시킨 후 16ms(256 points) 길이의 해밍 윈도우를 사용하여 5ms(80points)씩 쉬프트 시키면서 분석한다.

3.2 음소 인식 실험

화자 독립 음소 인식 실험에 있어서는 묵음을 제외한 47개의 유사 음소에 대하여 4상태 3출력의 Left-to-Right 모델을 구성하고 음성의 특징 파라미터로 MEL-LPC cepstrum과 LPC-MEL cepstrum을 이용하여 인식 실험을 수행하였다. 또한 이들 파라미터와 회귀계수를 결합한 경우에 대해서도 인식 실험을 수행하였다.

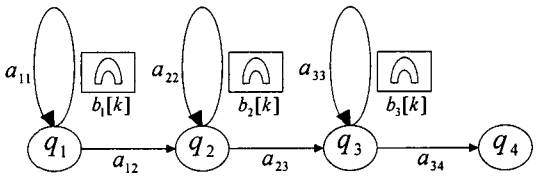

그림 3. HMM 음소 모델

음소 인식 실험은 LPC 분석을 하기 위한 Autocorrelation의 order와 최종 파라미터인 Mel-cepstrum의 order를 변화시키면서 인식률을 조사하였다.

다양한 인식 실험에서 MEL-LPC Cepstrum은 Autocorrelation order의 변화에 민감하지 않았지만 Mel-cepstrum order의 증가에 따라 인식률이 증가함을 알 수 있었다. 화자 독립 음소인식 실험결과를 표 1에 나타내었다.

표 1. 화자독립 음소인식률[%]

특징 파라미터	MEL-LPC	LPC-MEL
mel-cepstrum	51.08 (30-18)	47.61 (30-16)
mel + rgc	58.39 (30-20)	56.58 (30-10)

※괄호 안의 수치는 ((Auto. order)-(Mel Cep. order))임

표 1의 음소 인식 실험 결과에서 Mel-cepstrum만을 사용하였을 경우와 회귀계수를 함께 사용한 경우, MEL-LPC Cepstrum이 LPC-MEL Cepstrum을 사용한 경우보다 각각 약 3.5%, 1.8%정도 더 높은 인식률을 나타내었다.

3.3 단어 및 4연속 숫자음 인식 실험

앞 절의 음소 인식 실험 결과를 참고로 Autocorrelation order는 16차로, Mel-cepstrum order는 10차로 고정하여 단어 및 4연속 숫자음 인식 실험을 수행하였다. 인식의 기본 단위로는 48개의 유사음소단위(PLUs)를 음소모델로 사용하며, 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP법을 이용하였다.

먼저 445단어 인식 실험을 수행하고 각 특징파라미터에 따른 인식률의 변화를 표 2에 나타내었다.

표 2. 화자독립 단어인식률[%]

특징 파라미터	MEL-LPC	LPC-MEL
mel-cepstrum	80.97	78.13
mel + rgc	91.23	88.61

단어 인식 실험 결과, Mel-cepstrum만 사용한 경우와 Mel-cepstrum과 RGC를 사용한 두 경우 모두에서 MEL-LPC Cepstrum은 LPC-MEL Cepstrum을 사용한 경우보다 각각 약 2.8%, 2.6%정도 더 높은 인식률을 나타내었다. 다음으로 4연속 숫자음 인식 실험 결과를 표 3에 나타내었다.

표 3. 화자독립 4연 숫자음인식률[%]

특징 파라미터	MEL-LPC	LPC-MEL
mel-cepstrum	65.75	62.57
mel + rgc	76.86	72.86

4연속 숫자음 인식 실험 결과에서도 Mel-cepstrum만 사용한 경우와 Mel-cepstrum과 RGC를 사용한 두 경우 모두에서 MEL-LPC Cepstrum은 LPC-MEL Cepstrum을 사용한 경우보다 각각 약 3.1%, 4%정도 더 높은 인식률을 나타내었다.

이상의 결과로부터 원 음성신호에서 직접적으로 멜 주파수로 왜곡시킨 MEL-LPC Cepstrum이 음성의 특징 파라미터로 더 적합함을 알 수 있었다.

V. 결 론

본 논문에서는 한국어 음성인식 시스템의 성능 향상을 위해 청각 주파수 분해능을 가진 MEL-LPC cepstrum을 음소단위의 HMM(Hidden Markov Model)을 기반으로 하는 인식 시스템에 적용하여 그 결과를 비교 검토하였다.

음소단위의 HMM을 구성하기 위하여 음소인식에 적합한 Autocorrelation order차수와 Mel-cepstrum을 찾기 위한 실험과 이를 바탕으로 한 단어 및 연속 숫자음 인식 실험에서 MEL-LPC Cepstrum은 일반적으로 많이 사용되고 있는 LPC-MEL Cepstrum보다 높은 인식 결과를 나타내었다. 화자 독립 음성 인식 실험 결과로부터 MEL-LPC 분석 방법이 한국어 음성인식 시스템에서 효과적임을 확인하였다.

향후 이상의 결과를 바탕으로 대어휘 연속음성인식 시스템에 적용하고자 한다.

참고 문헌

[1] H. W. Strube, *Linear prediction on a warped frequency scale*, J.Acoust.Soc. America, 1988.

[2] Hiroshi Matsumoto, Yoshihisa Nakatoh, and Yoshinori Furuhata, *An Efficient MEL-LPC Analysis Method for Speech Recognition*, ICSLP'98, 1998

[3] Kazumasa Yamamoto and Seiichi Nakagawa, *Comparative Evaluation of Segmental Unit Input HMM and Conditional Density HMM*, ESCA. EUROSPEECH'95, 1994.

[4] X. D. Huang, Y. Ariki and M. A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh Univ., 1990.

[5] L. R. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, PTR Prentice-Hall, 1993.

[6] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *Spoken Language Processing*, PTR Prentice-Hall, 2001.