Modified-MFCC를 이용한 음성 특징 파라미터 추출 방법

이 상 복, 이 철 희, 정 성 환, 김 종 교
전북대학교 전자정보공학부
전화 : (063) 272-1177/ 팩스 : (063) 270-2400

Method of Speech Feature Parameter Extraction Using Modified-MFCC

Sang-Bok Lee, Chul-Hee Lee, Sung-Hwan Chung, Chong-Kyo Kim
Division of Electronics and Information Eng., Chonbuk National University
E-mail : sang960924@hanmail.net

Abstract

In speech recognition technology, the utterance of every talker have special resonant frequency according to shape of talker's lip and to the motion of tongue. And utterances are different according to each talker. Accordingly, we need the superior method of speech feature parameter extraction which reflect talker's characteristic well.

This paper suggests the modified-MFCC combined existing MFCC with gammatone filter. We experimented with speech data from telephone and then we obtained results of enhanced speech recognition rate which is higher than that of the other methods.

1. 서 론

음성(speech)은 일상 생활에서 인간은 개개인의 의사소통의 수단으로 이용하고 있다. 이러한 음성을 이용한 음성 인식 기술에서 음성 특징 파라미터를 추출하는 방법으로 현재는 LPC 계수로부터 유도된 LPC 켑스트럼과 인간의 청각 특성을 이용한 멜 주파수 켑스트럼(mel frequency cepstrum)계수를 이용한다. 또한, 청각 모델링(auditory modeling)을 이용한 파라미터 추출 방법 등이 있다.[1][3]

본 논문에서는 멜(mel) 단위 임계 대역(critical bandwidth)을 갖는 필터를 사용한 방법과 실제 청각 구조의 내이(inner ear)의 와우각(cochlear)에 있는 기저막(basilar membrane)의 특성이 gammatone 필터의 대역 통과 필터 열을 형성한다는 특성을 이용한 파라미터 추출 방법인 GFCC(gammtone filter frequency cepstrum coefficient)의 두 가지 방법을 변형한 형태의 음성 특징 파라미터를 추출하는 방법을 제안하였다.[2][3][4][5][7] 본 논문의 실험에서 전화 음성 DB를 이용하였다. 전화 음성의 채널왜곡이나 잡음에 대한 보상 기법이 있지만, 본 논문에서는 고려하지 않고 인식 실험을 하였다.

본 논문의 구성은 2절에서는 MFCC와, gammatone 필터를 이용한 GFCC에 대한 설명을 하고, 3절에서는 본 논문에서 제안한 modified-MFCC방법에 대해서 간단히 알아보도록 하겠다. 4절에서는 실험 및 결과, 5절에서 결론을 맺는다.

2. 음성 특징 추출 방법

2.1 MFCC(mel frequency cepstrum coefficient)

멜 주파수 켑스트럼 계수 (MFCC)는 현재 음성 인식에서 널리 사용되는 파라미터 추출 방법이다. 멜 단위(mel scale)는 Stevens과 Volkman(1940)에 의해 연구되어왔다.

$$\text{mel frequency} = 2595 \log_{10}\left(1+\frac{f}{700}\right) \quad (1)$$

그림 2.1은 임계 대역폭을 갖는 삼각 필터들을 나타냈다.

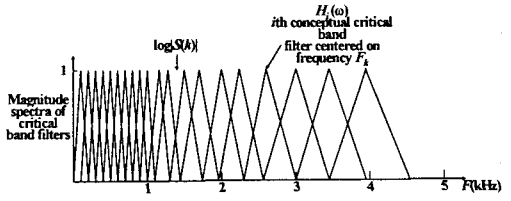

그림 2.1 멜(mel) 단위 임계 대역 필터

식(2)로 파라미터 계수 c_m을 구한다.

$$c_m = \frac{1}{M}\sum_{k=0}^{M-1} \tilde{Y}(k)\cos\left\{m\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right\} \quad (2)$$

여기서 M을 필터 수, m은 필터의 차수를 나타낸다.

2.2 GFCC(gammatone filter frequency cepstrum coefficient)

2.2.1 기저막 특성의 gammatone 필터

기저막은 앞쪽에서는 고주파에 민감하고 끝으로 갈수록 저주파에 민감하다. 이러한 특성으로 대역 통과 필터 열을 형성하고 모델화 하여 음성의 파라미터를 추출한다.

본 논문에서는 기저막 특성을 잘 묘사하는 gammatone 필터를 사용하였다. 이 필터는 기저막을 묘사하기 위해 많이 사용되는 필터로 차수 n이 4인 4차 gammatone 필터를 사용하였고 필터의 임펄스 응답으로 8차 recursive digital 필터로 구현할 수가 있다. 식 (3)은 gammatone 필터를 낸 것이다.

$$g(t) = \frac{at^{n-1}\cos(2\pi f_c t)}{e^{2\pi Bt}} \quad (3)$$

여기서 f_c는 필터의 중심 주파수, B는 f_c에서의 대역폭을 나타낸다. 대역폭 결정에 있어서 ERB (equivalent rectangular bandwidth)를 사용하였다.

$$ERB = \left[\left(\frac{f}{Q}\right)^{order} + B_n^{order}\right]^{\frac{1}{order}} \quad (4)$$

여기서 Q는 필터의 quality factor, B_n은 최소 대역폭을 나타낸다. 각 변수의 값은 여러 실험을 통해서 다양하게 제안되었다. 본 논문에서 사용한 값은 Glasberg 와 Moore 변수 값을 이용하였다. 또한 중심 주파수는 식 (5)와 같다.

$$f_{c_i} = -QB_n + (f_x + QB_n)e^{i(-\log(f_x+QBn)+\log(f_l+QB_n))/M}$$
$$i = 0, \cdots, M-1 \quad (5)$$

여기에서 f_x는 최대 주파수, f_l은 최소 주파수이고 M은 필터 개수를 나타낸다.

그림 2.2는 본 논문에서 사용한 40개의 gammatone 필터를 보인 것이다. 이들 필터는 Patterson 과 Holdworth 의 cochlea 설계 모델에서 기저막 특성의 ERB 대역폭을 갖는 gammatone 필터 뱅크로 구현된 필터 열들이다.[6]

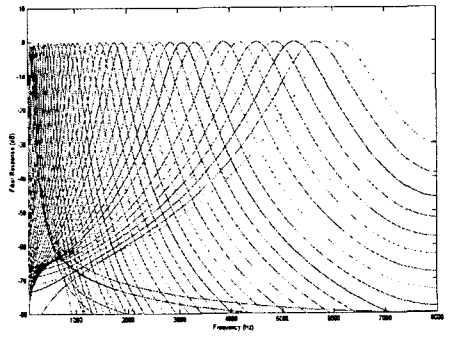

그림 2.2 ERB 대역폭을 갖는 gammatone 필터 뱅크

2.2.2 GFCC 파라미터 추출 방법

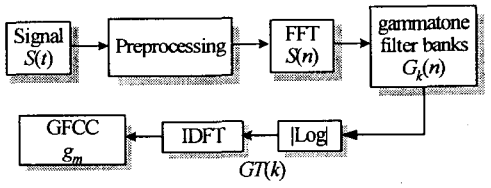

그림 2.3 특징 파라미터 추출 블록도

파라미터 계수를 구하는 방법은 그림 2.3의 블록도와 같이 MFCC를 구하는 방법과 유사하다. 기저막 특성

필터인 gammatone 필터를 적용하여 파라미터를 구하는 방법인 gammatone 필터 주파수 켑스트럼 계수 (GFCC : gammatone filter frequency cepstrum coefficients)이다. 따라서 멜 단위 임계 대역을 이용한 MFCC의 파라미터 보다 더 청각적인 특징을 가질 수 있게 된다.

$$GT(k)=\sum_{n=0}^{N/2}\log|S(n)|G_k(n),\quad k=1,\cdots,M \qquad (6)$$

여기서 M은 필터 수, N은 FFT 크기이다.

$$g_m=\frac{1}{M}\sum_{k=0}^{M-1}GT(k)\cos\left\{m\left(k+\frac{1}{2}\right)\frac{\pi}{M}\right\} \qquad (7)$$

식 (6)에서 보는 바와 같이 삼각 필터 대신에 gammatone 필터 $G_k(n)$를 삽입하여 파라미터 계수 g_m을 구하게 된다.

3. MMFCC(modified-MFCC)

본 논문에서는 2.2.1과 2.2.2에서 설명한 파라미터 추출 방법을 이용한 modified-MFCC 방법을 제안하였다. 제안한 방법은 음성 신호는 저주파 대역에서 그 특징이 잘 나타난다는 특성을 이용하였다. 따라서 저주파 대역의 음성 특징을 추출하기 위해 저주파 대역에서 해상도가 높은 gammatone 필터를 이용한 GFCC를 사용하여 약 1kHz이내에서 파라미터를 추출하고 그 이상에서는 MFCC를 이용하여 파라미터를 추출하도록 실험하였다. 그림 3.1은 파라미터 추출 블록도를 간단히 보인것이다.

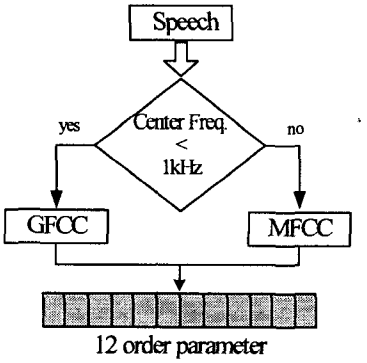

그림 3.1 파라미터 추출 블록도

4. 실험 및 결과

본 논문에서는 음성 인식 실험을 위해 전화 음성을 이용하였다. 본 음성은 전북대학교 음성처리연구실의 프로젝트 수행과정 중 수집된 전화음성으로 전라도, 충청도, 경상도등 각 지역에서 유선전화와 무선전화로 수집된 음성 데이터이다. 본 논문에 사용된 음성 데이터의 환경은 다음과 같다.

- 전화음성 데이터
  - 남성화자 : 28명
  - 여성화자 : 28명
  - 단어 수 : 10개
  - 발성회수 : 1회
  - 총 : 560개
    - 학습 데이터: 남성화자- 20명, 여성화자- 20명
    - 시험 데이터: 남성화자- 8명, 여성화자- 8명
- 수집 환경
  - 주변잡음이 있는 장소에서 수집
  - 8kHz의 16 bit 데이터

그림 4.1과 4.2는 전화 음성과 마이크 음성의 파형과 스펙트로그램으로 나타낸 것이다. 전화음성이 심한 잡음이 있음을 알 수 있다.

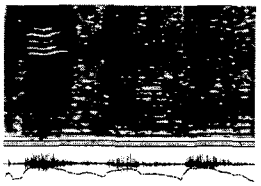

그림 5.1 전화음성

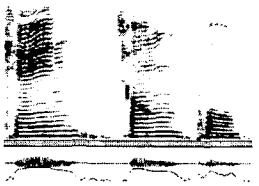

그림4.2 마이크음성

실험은 분석 프레임의 길이는 25.625ms, 이동 프레임의 크기는 12.5ms로 하여 각각의 LPCC, MFCC, GFCC와 MMFCC의 12차의 파라미터를 추출하여 인식률을 비교하였고, 전화 음성에 대한 보상은 고려하지 않고 실험하였다. 인식 알고리즘으로는 LBG 알고리즘을 이용한 64개의 코드북을 만들어 상태 수 8개인 이산HMM을 사용하여 수행하였다. 표 1은 기존의 LPCC와 MFCC, GFCC와 제안한 방법, MMFCC의 전체 인식률 값이고, 표 2는 각각 10개의 단어에 대한 인식률 값을 나타냈다.

표 1. 전체 인식률 값

	MMFCC	GFCC	MFCC	LPCC
인식률 (%)	90	82.5	82.5	72.5

표 2. 각각 단어의 인식률

단어	MMFCC	GFCC	MFCC	LPCC
1	93.75	81.25	93.75	62.50
2	87.50	93.75	93.75	100.00
3	87.50	93.75	81.25	93.75
4	87.50	81.25	62.50	68.75
5	93.75	75.00	81.25	62.50
6	93.75	75.00	81.25	75.00
7	93.75	75.00	68.75	68.75
8	87.50	87.50	93.75	75.00
9	87.50	81.25	81.25	50.00
10	87.50	81.25	87.50	68.75

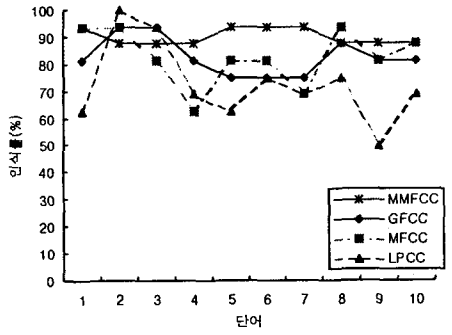

그림 4.3 각각 단어의 인식률 그래프

그림 4.3과 표2에서 보는 바와 같이 GFCC와 MFCC의 전체 인식률 값은 같지만 각각의 단어에 대한 인식률 값은 GFCC가 일정값 75%이상의 균등한 값을 나타냈다. 또한 MMFCC는 두가지 파라미터의 특성을 반영해서 평균 90%인 인식률과 각각의 단어에 대해서 87%이상의 인식값을 나타냈다.

5. 결 론

음성 특징 파라미터 추출 방법들 중에 MFCC는 인간의 청각 특성을 반영한 추출법으로 많이 이용되고 있다. MFCC의 청각 특성은 1kHz이하에서는 선형적이고 이상에서는 지수적인 값을 갖는 멜 크기의 대역 폭을 갖는 삼각 대역 필터를 이용하는데 그쳤다. 하지만 본 논문에서는 실제 청각 모델링(auditory modeling)에서 파라미터 추출에 이용하는 기저막(basilar membrane)의 특성인 gammatone 대역 통과 필터를 MFCC를 구하는 방법 중 필터 뱅크의 부분에 대치함으로써 파라미터 추출 방법을 고안했고, MFCC와 GFCC의 방법을 결합한 형태의 파라미터 추출 방법을 제안하여 실험하였다. 실제 청각 모델링(auditory modeling)을 하지 않고 청각적 특성이 가장 가까운 특징을 가진 파라미터를 추출할 수 가 있다.

실험 결과 변형된 MFCC가 다른 파라미터 추출방법에 비해 평균 8%이상의 높은 인식률 값을 나타냈다.

실험은 전화 음성에 대한 보상을 고려하지 않고 하였는데, 보상을 고려한 실험이 또한 필요하고, 인식률을 향상하기 위한 훈련 기법의 보완과 다른 잡음 음성 DB에 대한 더 많은 인식 실험이 필요하다.

참고문헌

[1] Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice- Hall, 1993.

[2] 정호영, 김도영, 은종관, 이수영,"청각구조를 이용한 잡음 음성의 인식 성능 향상", 한국음향학회지 제14권, 제5호, 1995.

[3] John R. Deller, Jr., John G. Proakis, John H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.

[4] J. M. Kates, "A Time-Domain Digital Cochlear Model," *IEEE Trans. on Signal Processing*, vol. 39, no. 12, pp. 2573-2592, Dec. 1991.

[5] Rivarol Vergin, Douglas O'Shaughnessy, "Generalized Mel Frequency Cepstral Coefficients for Large-Vocabulary Speaker-Independent Con- tinuous-Speech Recognition," *IEEE Trans. on Speech & Audio Processing*, vol. 7, no. 5, pp. 512-532, 1999.

[6] M. Slaney, "An Efficient Implementation of the Patterson-Holdworth Auditory Filter Bank," *Apple Computer Tech. Report #35*, 1993.

[7] 김종교, 이철희, 신유식,"전화망 음성 인식을 위한 Gammatone Filter 특징 추출," 전북대학교 공학연구 논문집, vol. 31, pp. 107-113, 2000. 12.