

남녀의 음향학적 특징벡터의 비교 분석에 관한 연구

최재승*, 정병구**

*신라대학교 전자공학과, **목포대학교 전기공학과

E-mail : *jschoi@silla.ac.kr, **jbg8917@hanmail.net

요 약

본 논문에서는 캡스트럼 계수의 변화에 따른 남성화자와 여성화자의 음향학적인 특징벡터를 비교하여 분석하는 기초적인 연구를 수행한다. 특히 FFT 캡스트럼 및 LPC 캡스트럼에 대한 남녀의 음향학적인 특징벡터의 차이점을 나타낸다. 향후 이러한 차이점을 기초로 하여 신경회로망 등에 의한 성별 인식에 대한 연구를 수행함으로써 남성화자 및 여성화자를 분리할 수 있는 근거를 마련하는 기초연구이다.

키워드

캡스트럼 계수, 남녀성별인식, 특징벡터

I. 서 론

근년 정보통신 및 컴퓨터기술이 급속하게 발전하게 되면서 특히 음성에 의한 단말기 시스템 기능의 성능이 고조되어, 이러한 음성 기술의 실현을 위한 음성인식기술의 연구개발 및 노력이 활발하게 수행되고 있다 [1-3]. 이러한 음성인식 기술 중에서 성별 화자인식 기술은 남녀 화자를 대상으로 하여 성별 인식실험을 하게 되는데, 일반적으로 신경회로망(Neural Network)의 연구가 보고되고 있다[4]. 신경회로망은 1960년대부터 패턴인식에 응용하려는 연구가 활발히 진행되어 왔으며, 또한 오차역전파 학습 알고리즘의 등장으로 인하여 여러 분야에서 응용하게 되었다[5]. 이 오차역전파 학습 알고리즘은 계산기 상에서 실현하기 쉽게 구현한 것이며, 다방면의 연구자가 각각의 응용모델을 구성하여 신경회로망의 응용이 활발해지게 되었다.

본 논문에서는 남녀성별 인식을 위하여 남성 및 여성화자의 특징벡터를 비교분석하여, 이러한 음향학적인 특징벡터의 차이점을 비교 분석한다. 본 논문에서는 남녀의 특징벡터를 추출한 후에 비교분석 과정을 수행함으로써 향후에 각 화자에 대한 성별을 신경회로망 등의 패턴인식에 의하여 구별되도록 하는 음성인식에 관한 연구의 기초를 마련하고자 한다.

II. 음성의 특징벡터 추출

선형예측부호화(Linear Predictive Coding, LPC) 분석법은 음성분석 및 합성 등의 여러 분야에서 사용되는 기법이며, 음성스펙트럼이 가진 특성을 상대적으로 적은 수의 파라미터만으로 정확하게 표현할 수 있다.

이산적인 시간 n 에서의 음성 샘플을 x_n 이라고 하자. 이때, 시간 n 에서의 음성 샘플의 예측값 \hat{x}_n 이라고 하면, 식 (1)처럼 표현될 수 있다.

$$\hat{x}_n = \sum_{k=1}^p a_k x_{n-k} = a_1 x_{n-1} + \dots + a_p x_{n-p} \dots (1)$$

따라서 음성신호의 현재 값은 식 (1)로부터 이전 p 개의 과거의 값 x_{n-1}, \dots, x_{n-p} 로부터 예측된다. 따라서 실제 입력된 값과 예측된 값 간의 차이를 나타내는 선형예측오차를 $e(n)$ 이라고 하면 식 (2)가 성립한다.

$$e_n = x_n - \hat{x}_n = x_n - \sum_{k=1}^p a_k x_{n-k} \dots (2)$$

여기에서 $a_k(k=1, \dots, p)$ 를 선형예측계수 혹은 LPC 계수라고 하며 e_n 를 잔차신호라고 한다. 따라서 선형예측오차 e_n 의 2승 평균값이 최소가 되도록 LPC 계

수를 구한다. 본 논문에서는 저역에 음성신호가 주로 분포하고 있다는 특징을 이용하여, 저역 주파수영역에서의 유성음이 가지고 있는 특유의 주파수 스펙트럼 구조를 채용한 LPC 켈스트럼 계수[6]의 특징벡터를 사용한다. LPC 켈스트럼 계수는 스펙트럼 포락의 모양을 나타내는 정보로서, 음성신호를 표현하기 위한 중요한 수단으로서 사용된다.

고속 푸리에 변환(Fast Fourier Transform, FFT)에 의한 FFT 켈스트럼 방법은 스펙트럼의 포락성분만을 추출하여 평균화된 스펙트럼 성분을 구하는 방법으로, 스펙트럼 대수의 척도에 의하여 구해지는 추정방법이다. FFT 켈스트럼이 가진 큰 특징은 낮은 케프런시 대역으로부터 로그가 취해진 스펙트럼 포락 정보와 높은 케프런시 대역에서 나타나는 스펙트럼의 주기성을 표현하는 세부구조 정보를 분리해 낼 수 있다. 따라서 본 논문에서는 제안하는 켈스트럼 방법은 시간영역에서 잡음이 중첩된 음성신호를 FFT에 의해서 구해지는 FFT에 의한 켈스트럼[7]에 대해서 기술한다. 이 FFT 켈스트럼 방법은 음성신호를 FFT, 대수연산 조작, 역 푸리에 변환, 켈스트럼에 창 등의 처리를 하여, 스펙트럼의 포락 정보에 해당하는 저역의 FFT 켈스트럼 계수를 구한다.

III. 음성특징벡터의 비교 실험

본 논문에서는 신경회로망의 입력으로 사용하기 위하여 LPC 켈스트럼 계수, FFT 켈스트럼 및 실효값(Root Mean Square, RMS), LPC 켈스트럼 및 FFT 스펙트럼 등의 음성특징 벡터를 사용한다. 그림 1은 남성화자에 대한 10차의 LPC 켈스트럼 계수의 변화 모양을 나타내며, 그림 2는 여성화자에 대한 10차의 LPC 켈스트럼 계수의 변화모양을 나타낸다. 그림 3과 그림 4는 각각 남성화자 및 여성화자에 대한 12차의 LPC 켈스트럼 계수의 변화의 모양을 나타내고 있다. 그림으로부터 알 수 있듯이 그림 1과 그림 3의 남성화자에 대해서는 LPC 켈스트럼의 모양이 비교적 완만하게 변화하고 있는 반면에 그림 2와 그림 4의 여성화자에 대해서는 굴곡이 많은 변화의 형태를 하고 있는 것을 알 수 있다. 따라서 이러한 남성 및 여성화자 간의 차이점에 의해서 향후의 연구에 사용할 신경회로망 등의 학습에 의하여 충분히 식별이 가능할 것으로 판단된다. 그림 5와 그림 6은 각각 남성 및 여성화자에 대하여 12차의 LPC 켈스트럼 계수와 8차의 FFT 스펙트럼에

대한 계수의 변화 모양을 나타내고 있다. 그림에서 가로축의 1차부터 12차의 계수가 LPC 켈스트럼 계수(전반부)를 나타내며 13차부터 20차까지가 FFT 스펙트럼(후반부)을 나타낸다. 그림 1, 2, 3, 4의 남성 및 여성화자의 모양과 비슷하게 변화하는 것을 알 수 있으며, 그림 5와 그림 6의 후반부 FFT 스펙트럼에 대해서도 남성화자 및 여성화자에 대한 분명한 차이점을 확인할 수 있다.

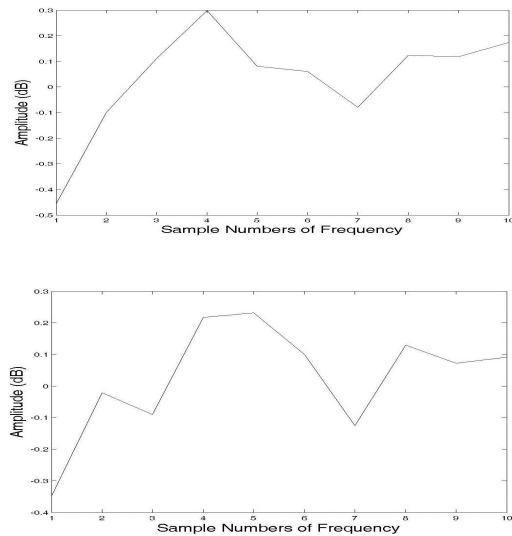


그림 1. 10차 LPC 켈스트럼 계수의 변화(남성화자)

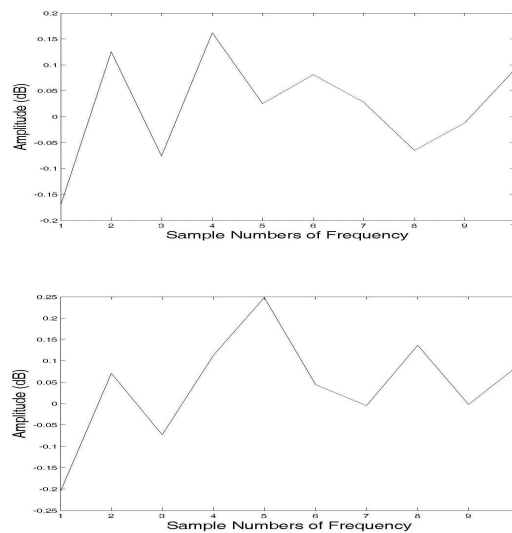


그림 2. 10차 LPC 켈스트럼 계수의 변화(여성화자)

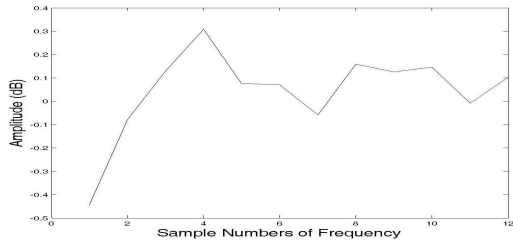


그림 3. 12차 LPC 켈스트럼 계수의 변화(남성화자)

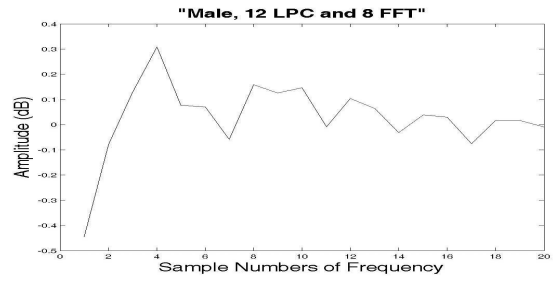
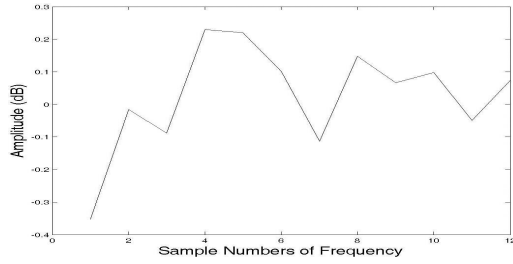


그림 5. LPC 켈스트럼 계수 및 FFT 스펙트럼의 변화(남성화자)

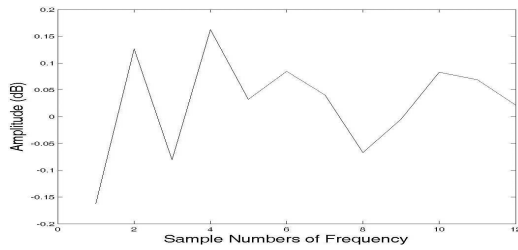
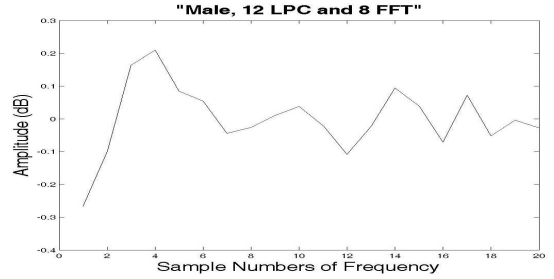


그림 4. 12차 LPC 켈스트럼 계수의 변화(여성화자)

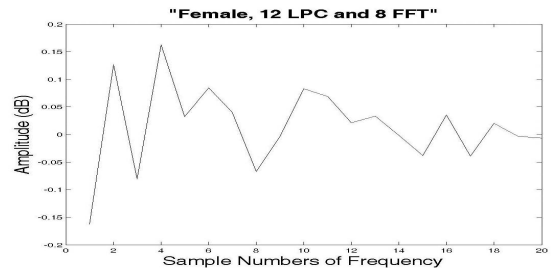
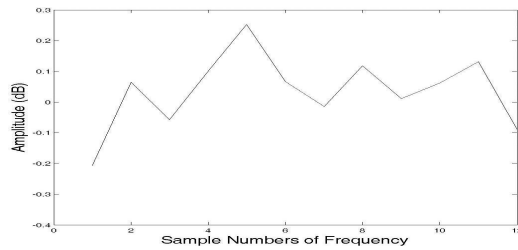
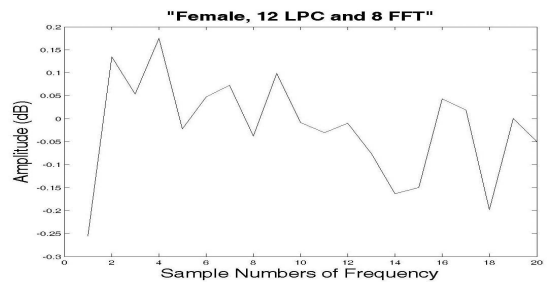


그림 6. LPC 켈스트럼 계수 및 FFT 스펙트럼의 변화(여성화자)



IV. 결론

본 논문에서는 남성화자와 여성화자의 특징벡터를 비교 분석하였다. 특히 LPC 켈스트럼 및 FFT 켈스트럼, 실효값에 대한 남녀의 음향학적인 특징벡터의 차이점을 비교 분석하여 명확한 차이점을 확인할 수 있

었다. 향후 이러한 차이점을 이용하여 신경회로망 등에 의한 남녀성별 인식에 대한 연구를 수행함으로써 남성화자 및 여성화자를 충분히 분리할 수 있는 기초를 마련했다고 본다. 그러나 향후의 연구과제로서는 실험 데이터를 충분히 확보하여 실험할 필요가 있다.

참고문헌

- [1] J. G. van Velden and G. F. Smoorenburg, "Vowel recognition in noise for male, female and child voices", International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 989-992, 1991.
- [2] F. Hassan, M.R.A. Kotwal, and M.N. Huda, "Bangla ASR design by suppressing gender factor with gender-independent and gender-based HMM classifiers", World Congress on Information and Communication Technologies, pp. 1276-1281, 2011.
- [3] A. A. M. Abushariah, T. S. Gunawan, O. O. Khalifa, and A. A. M. Abushariah, "English digits speech recognition system based on Hidden Markov Models", International Conference on Computer and Communication Engineering, pp. 1-5, 2010.
- [4] Y. Konig and N. Morgan, "GDNN: a gender-dependent neural network for continuous speech recognition", International Joint Conference on Neural Networks, Vol. 2, pp. 332-337, 1992.
- [5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagation errors", Nature, Vol. 323, pp. 533-536, 1986.
- [6] P. B. Patil, "Multilayered network for LPC based speech recognition", IEEE Transactions on Consumer Electronics, Vol. 44, No. 2, pp. 435-438, 1998.
- [7] J. He, L. Liu, and G. Palm, "On the use of residual cepstrum in speech recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 5-8, 1996.
- [8] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", in Proc. ISCA ITRW ASR2000 on Automatic Speech Recognition: Challenges for the Next Millennium, Paris, France, 2000.