LPC 켑스트럼 및 FFT 스펙트럼에 의한 성별 인식 알고리즘

최재승*, 정병구**

*신라대학교 전자공학과, **목포대학교 전기공학과

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

요 약

본 논문에서는 입력된 음성이 남성화자인지 여성화자인지를 구분하는FFT 스펙트럼 및 LPC 켑스트럼 입력에 의한 성별인식 알고리즘을 제안한다. 본 논문에서는 특히 남성화자와 여성화자의 특징벡터를 비교 분석하여, 이러한 남녀의 음향학적인 특징벡터의 차이점을 이용하여 신경회로망에 의한 성별 인식에 대한 실험을 수행한다. 특히 12차의 LPC 켑스트럼 및 8차의 저역 FFT 스펙트럼의 특징벡터를 사용한 경우에, 남성화자 및 여성화자에 대해서 양호한 남녀 성별인식률이 구해졌다.

키워드

남녀성별인식, 인식시스템, 신경회로망, 오차역전파 알고리즘

Ⅰ. 서 론

최근에 음성신호처리 기술의 실현을 위한 음성인식 기술 중에서 특히 남성화자 및 여성화자를 모델로 한 성별 화자인식에 대한 연구개발도 보고되고 있다[1]. 이러한 성별 화자인식 기술에 대한 연구 논문으로 은닉 마르코프 모델(Hidden Markov Model, HMM), 신경회로망(Neural Network) 등의 연구가 보고되고 있다[2, 3]. 이 중에서 신경회로망은 오래전부터 패턴인식에 응용하려는 연구가 활발히 진행되어 왔으며, 또한 오차역전파 학습 알고리즘의 등장으로 인하여 여러 분야에서 응용하게 되었다[4]. 이 오차역전파 학습 알고리즘은 계산기 상에서 실현하기 쉽게 구현한 것이며, 다방면의 연구자가 각각의 응용모델을 구성하여 신경회로망의 응용이 활발해지게 되었다. 음성의 분야에서도 음성인식, 화자인식 등의 여러 분야에서 3층의 신경회로망과 오차역전파 학습 알고리즘을 사용하여 음성분류가 가능하게 되었다.

따라서 본 논문에서는 이러한 신경회로망의 오차역전파 학습 알고리즘에 기초한 남녀 성별인식 알고리즘을 제안한다. 본 논문에서는 특히 남성 및 여성화자의 특징벡터를 사용하여 신경회로망에 의한 남녀 성별인식에 대한 실험을 수행한다. 제안하는 알고리즘에서는 남녀의 특징벡터를 추출하여 각 화자에 대한 성별을 신경회로망의 패턴인식에 의하여 구별하도록 한다. 이러한 패턴인식 방법을 이용하여, 본 논문에서는 기초적인 화자종속 음성인식에 의하여 각 화자의 성별을 구별하는 실험을 하여, 일반적인 종래의 방법들과 비교하는 실험을 수행한다. 따라서 본 논문에서는 이러한 연구배경에 기초하여 음성의 특징벡터로는 선형예측에 의한 켑스트럼 계수 및 푸리에 변환에 의한 켑스트럼 계수 등의 특징벡터를 사용한다.

Ⅱ. 남녀성별 인식 알고리즘

본 논문에서는 새롭게 신경회로망에 입력된 음성이 남성화자인지 여성화자인지를 판별하기 위하여 출력층의 유닛을 2개로 설정한다. 신경회로망의 학습 계수는 α=0.1, 가속도 계수는 β=0.03으로 하였으며, 최대 학습

횟수는 10,000회로 하였다.

그림 1은 본 논문에서 제안한 성별인식 알고리즘의 블록도이다. 먼저 음성신호의 한 프레임을 256샘플(32 ms)로 분리한 후에 해밍창을 통과시킨다. 그리고 신경회로망의 입력신호로 사용하기 위하여 1) 12차의 저역의 LPC(Linear Predictive Coding) 켑스트럼 계수(10 입력 유닛) 및 8차의 저역 FFT 스펙트럼(20 입력 유닛)을 신경회로망에 입력하여 남성화자 및 여성화자로 인식되도록 3층 구조의 신경회로망에 의해서 학습된다. 새로운 화자가 입력되었을 때 신경회로망에 의해서 출력된 가중치를 사용하여 남성화자 및 여성화자로 최종적으로 인식된다.

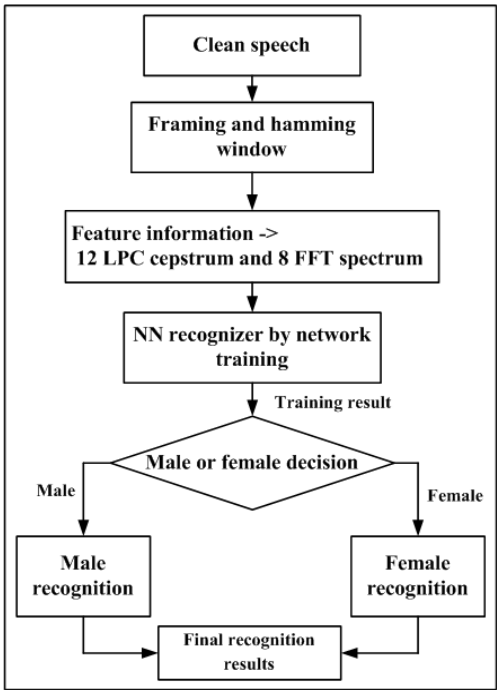

그림 1. 제안한 남녀성별인식 알고리즘

III. 실험 결과

본 실험에서 사용한 음성신호는 8 kHz의 샘플링 주파수를 가진 환경에서 녹음된 영어숫자로 구성된 Aurora2 데이터베이스(Database, DB)[5]를 사용하였다. Aurora2 DB의 모든 음성데이터는 남성화자 55명 및 여성화자 55명에 의해서 발성된 음성을 녹음한 총 8440개의 숫자로 된 테스트 셋 A, B, C의 음성데이터로 구성되어 있다. 제안한 시스템은 Aurora2 DB 중에서 테스트 셋 A의 잡음이 중첩되지 않은 음성데이터에서 남성화자 및 여성화자를 임의로 선택하여 화자인식 실험을 수행하며, 화자인식률에 의하여 인식 성능을 평가한다. 본 논문에서의 화자 인식률의 평가는 입력음성의 전체 프레임에 대하여 각 프레임에서 신경회로망의 출력값이 정확하게 인식되는 프레임 비율로 정의한다. 사용한 학습 및 테스트 데이터는 4명의 화자(Speaker A, B, C, D)의 문장을 사용하였다. 음성 데이터는 256 샘플에서 약 20 프레임부터 70 프레임 정도를 나누워진다.

표 1은 본 논문에서 제안하는 방식으로, 신경회로망의 입력데이터로서 12차의 LPC 켑스트럼 및 8차의 FFT 스펙트럼을 입력하였을 때(20-20-2 네트워크)의 성별 인식률을 나타낸다. 표 5에서 남성화자에 대한 인식률이 평균 99.8%이며, 이 때의 오인식률의 평균은 0.2%로 나타났다. 여성화자에 대해서는 인식률이 평균 96.5%이며, 오인식률은 남성화자에 비해서 다소 높은 3.5%로 나타난 것을 알 수 있다.

표 1. 12차의 LPC 켑스트럼 및 8차의 FFT 스펙트럼에 대한 성별 인식률

Speaker	Recognition rates[%]			
	Male / (ER)		Female / (ER)	
A	100.0%	(0.0%)	94.4%	(5.6%)
B	99.3%	(0.7%)	100.0%	(0.0%)
C	100.0%	(0.0%)	100.0%	(0.0%)
D	100.0%	(0.0%)	91.5%	(10.5%)
Average	99.8%	(0.2%)	96.5%	(3.5%)

IV. 결론

본 논문에서는 음성장치에 음성이 입력될 때에 입력된 음성이 남성화자인지 여성화자인지를 구분하는 성별인식 알고리즘을 제안하였다. 본 논문에서는 남성화자와 여성화자의 특징벡터를 이용하여 신경회로망에 의한 성별 인식에 대한 실험을 수행하였다. 12차의 LPC 켑스트럼 및 8차의 저역 FFT 스펙트럼의 특징벡터를 사용한 경우에, 남성화자에 대해서는 평균 99.8%, 여성화자에 대해서는 평균 96.5%의 남녀 성별인식률이 구해졌다. 따라서 본 실험에서 비교한 다른 방법 및 기존 방법들보다 우수하다는 것을 실험으로 확인할 수 있었다. 그러나 향후의 연구과제로서는 학습에 사용하지 않은 데이터를 사용하여 좀 더 많은 문장을 사용한 남녀 성별 화자독립 인식 알고리즘을 개발할 필요가 있다. 또한 잡음이 중첩된 음성에 대해서도 연구의 검토가 필요하다고 본다.

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] 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.

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagation errors", Nature, Vol. 323, pp. 533-536, 1986.

[5] 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.