켑스트럼 계수에 의한 모음검출을 위한 음성인식

최재승*

*신라대학교 전자공학과

Speech Recognition for Vowel Detection using by Cepstrum Coefficients

Jae-Seung Choi*

*Department of Electronic Engineering, Silla University

E-mail : *jschoi@silla.ac.kr

요 약

본 논문에서는 켑스트럼 계수를 이용하여 음성인식을 하는 알고리즘을 제안한다. 본 논문에서 제안하는 방법은 사람이 발성한 음성을 두 영역의 켑스트럼 계수로 분리한 후에, 신경회로망을 사용하여 음성인식을 하는 방법이다. 본 논문에서 제안하는 신경회로망은 오차가 거의 없어지는 일정 기간 동안 네트워크를 학습시킨 후에 신경회로망의 학습 데이터와는 다른 새로운 음성이 신경회로망에 입력된 경우에 대하여 각 음성 구간에서 분류가 가능한 모음검출을 위한 음성인식 시스템을 제안한다.

키워드

Speech recognition, neural network, cepstrum coefficients, recognition rate.

Ⅰ. 서 론

최근에 인간의 음성은 인간과 컴퓨터와의 상호대화를 위한 매개체로서의 역할이 중요하게 되었다. 음성은 모든 인간들이 이용할 수 있는 기본적인 정보통신 수단이며, 수많은 정보 전달 방법 중에서도 음성은 가장 보편적인 수단이라고 할 수 있다. 이러한 음성을 하나의 정보통신 수단으로 하여 음성인식 및 화자인식 위하여 신경회로망(Neural Network, NN)을 이용한 컴퓨터 시뮬레이션을 통한 활발한 연구가 수행되고 있다[1-5].

신경회로망을 이용한 음성인식 방법에서는 신경회로망은 학습과정을 통해 입력층에서 중간층 혹은 중간층에서 출력층으로 향하는 뉴런(neuron)들 간을 연결하는 가중치(weight)를 변경하여 오차를 최소화하게 함으로써 새로운 입력데이터에 대하여 분류가 가능하도록 음성인식 과정을 수행한다. 본 논문에서는 음성 인식에 대하여 성능이 우수한 오차 역전파 학습 알고리즘을 이용한 신경회로망을 사용하여 음성 인식을 수행한다. 일반적으로 신경회로망은 외부로부터 입력되는 음성의 특징 데이터를 추출하여 신경회로망의 네트워크의 학습 과정을 통하여 그 특징을 분류할 수 있는 특성을 가지고 있다. 따라서 신경회로망은 각 음성의 입력 특징 데이터의 차이에 의하여 음성인식 분류에 대한 오류를 최소화할 수 있기 때문에, 뛰어난 패턴 인식 능력 처리 구조로 인하여 숫자음이나 연속음 등과 같은 목적에 적합한 음성인식 방법이라 할 수 있다[1-5].

본 논문에서는 신경회로망을 이용하여 여러 사람이 발성한 음성을 입력하여 각 개인이 가지고 있는 음성의 특징을 추출한 후에 이 특징 입력데이터를 신경회로망의 입력값으로 한다. 신경회로망을 오차가 거의 없어지는 일정 기간 동안 네트워크를 학습시킨 후에 신경회로망의 학습 데이터와는 다른 새로운 음성이 신경회로망에 입력할 경우에 대하여 각 음성을 인식하고 판단할 수 있는 모음검출을 위한 음성인식 시스템을 제안한다.

II. 켑스트럼 계수 분석

본 논문에서는 고속 푸리에 변환(Fast Fourier Transform, FFT)에 의해서 구해지는 FFT에 의한 켑스트럼을 분석한다.

FFT 켑스트럼 방법은 스펙트럼 대수의 척도에 의하여 구해지는 스펙트럼 포락에 의한 추정방법이며, 켑스트럼에 창(window)를 씌움으로써 음원의 주기성에 해당하는 성분을 제거하여, 스펙트럼 포락성분의 단시간 영역성분 만을 추출함으로써 평균화된 스펙트럼 성분을 구하는 방법이다. 본 실험에서는 샘플링 주파수 8 kHz의 이산시간신호를 128샘플(16 ms)의 프레임으로 분리하여 각 프레임의 샘플값을 해밍창를 통과시킨 후에 켑스트럼 변환(FFT→log| |→IFFT)을 한다. 구해진 켑스트럼을 켑스트럼창에 통과시킴으로써 켑스트럼의 저역부의 12개의 켑스트럼 데이터를 구한다.

본 실험에서 사용한 음성신호는 8 kHz의 샘플링 주파수를 가진 환경에서 녹음된 연결된 영어 숫자로 구성된 Aurora2 데이터베이스(Database, DB)[6]를 사용하였다. Aurora2 DB의 모든 음성데이터는 ETSI (European Telecommunications Standards Institute)로부터 배포되었으며, 테스트 셋 A, B, C의 음성데이터로 구성되어 있다. Aurora2 데이터베이스는 남성화자 55명 및 여성화자 55명에 의해서 발성된 음성을 녹음한 총 8440개의 숫자로 구성된 테스트 셋 A, B, C의 음성데이터를 사용하였다.

III. 제안한 인식 알고리즘

본 논문에서는 음성을 사용한 음성인식에 신경회로망을 사용한 음성인식의 식별방법을 제안한다. 본 논문에서 제안하는 음성인식 알고리즘의 처리 과정을 그림 1과 같이 나타낸다. 제안하는 음성인식 과정은 크게 음성신호의 전처리 과정, FFT 과정, 켑스트럼 추출 과정, 정규화 과정, 신경회로망에 의한 분류 과정 등의 단계로 분류할 수 있다.

그림 1의 음성인식 알고리즘에서 발성된 입력 음성에 대하여 입력된 이산시간 신호를 128샘플의 프레임으로 분리하여 각 프레임의 샘플값을 해밍창을 통과시킨다. 이 후에 FFT 변환을 한 후에 FFT 켑스트럼 계수를 추출한다. 그리고 추출된 FFT 켑스트럼 계수를 정규화한다. 정규화된 켑스트럼 계수를 신경회로망에 입력하여 모음부를 인식한다.

따라서 제안한 음성인식 시스템에서의 신경회로망은 각각 입력층의 유닛수는 12개의 켑스트럼을 신경회로망에의 입력으로 한다. 중간층 유닛은 20개, 출력층은 1개의 유닛으로 구성된다. 따라서 학습을 통해 기억된 각 가중치들을 사용하여 새로운 화자의 음성이 입력될 경우 기억된 가중치들을 가지고 새로운 각 화자에 해당하는 음성에 대하여 유성음인가 아닌가를 인식하게 된다.

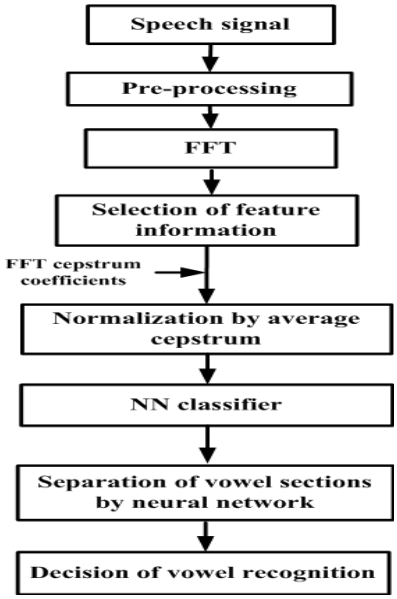

그림 1. 제안한 음성인식 알고리즘

IV. 실험결과

본 논문에서 제안한 시스템은 임의적으로 각 음성에 의한 특정 단어를 데이터베이스로부터 선택하여 음성인식 실험을 수행하여 음성인식률에 의하여 인식 성능을 평가한다.

실험 수행 시 각각의 실험에 대하여 충분한 검증을 위하여 각각의 인식방법에 대해 총 10번 실험결과의 평균치를 사용하여 인식률을 산출하였다. 사용한 학습데이터는 20문장을 사용하였으며 학습으로 10개를, 테스트로 10개를 사용하였다.

신경회로망에 입력되는 음성의 특징 벡터로는 유성음 구간에 대해서는 12차 FFT Cepstrum 계수를 사용하여, 유성음에 대하여 학습 시에는 평균 91.7%, 테스트 시에는 평균 89.3%의 인식율을 구하였다.

V. 결론

본 논문에서는 유성음 구간에 대한 음성인식의 성능개선을 위하여 신경회로망을 사용하여 모음의 음성 인식률을 향상시키는 방법을 제안하였다. 제안한 신경회로망은 음성인식의 성능을 개선하기 위하여 오차 역전파 학습 알고리즘을 이용하여 네트워크를 학습시켰다. 제안한 음성인식 알고

리즘은 발성음성의 유성음 구간을 검출하여, 특징 데이터를 추출한 후 이 특징데이터를 신경회로망에 적용시켜 모음 음성을 인식하는 방법을 적용하였다. 제안한 인식방법은 실험을 통하여 인식성능을 확인하였다. 향후 연구 과제로는 좀 더 많은 어휘의 인식이 가능한 음성인식 알고리즘을 연구할 예정이다.

참고문헌

[1] T. T. Le, J. S. Mason and T. Kitamura, "Characteristics of multi-layer perceptron models in enhancing degraded speech", Proc. ICSLP-94, pp. 1611-1614, 1994.

[2] S. K. Pal, S. Mitra, “Multilayer perceptron, fuzzy sets, and classification”, IEEE Transaction on Neural Networks, Vol. 3, No. 5, pp. 683-697, 1992.

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

[4] A. V. Ooyen and B. Nienhuis, "Improving the convergence of the back-propagation algorithm," Neural Networks 5, 3, pp. 465-471, 1992.

[5] W. G. Knecht, M. E. Schenkel, G. S. Moschytz, “Neural network filters for speech enhancement”, IEEE Trans. Speech and Audio Processing, Vol. 3, No. 6, pp. 433-438, 1995.

[6] 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.