다층 퍼셉트론 네트워크에 의한 연속음성 화자분류

최재승*

*신라대학교

Jae-Seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

주변의 배경잡음으로부터 음성인식률을 향상시키기 위하여 적절한 음성의 특징 파라미터를 선택하는 것이 매우 중요하다. 본 논문에서는 위너필터 방법이 적용된 인간의 청각 특성을 이용한 멜 주파수 켑스트럼 계수를 사용한다. 제안한 멜 주파수 켑스트럼 계수의 특징 파라미터를 다층 퍼셉트론 네트워크에 입력하여 학습시킴으로써 화자인식을 구현한다.

키워드

위너필터, 멜 주파수 켑스트럼, 배경잡음, 화자인식

Ⅰ. 서 론

근래 음성인식 기술의 발전과 더불어 현대 사회에서 실용화가 가능한 음성대화 시스템을 염두에 둔 음성인식 시스템 개발이 진행되어 오고 있다[1]. 이러한 음성대화 시스템을 구축할 때 고려해야 할 사항으로는 화자의 정확한 분류, 주변잡음에 의한 음성시스템의 오동작 검증, 음성구간의 검출 등이 있다[2].

실제 환경에서 기인하는 주변 배경잡음 때문에 음성인식률이 떨어지는 문제점도 발생되고 있다. 따라서 다양한 배경잡음에 대한 대처방안으로서 음성인식 시스템의 성능을 향상시키기 위하여 음성강조, 화자인식, 위너필터 방법 등이 제안되고 있다[3, 4].

본 논문에서는 위너필터를 적용한 멜 주파수 켑스트럼 계수(Mel Frequency Cepstrum Coefficient, MFCC)[5]의 파라미터를 이용한 다층 퍼셉트론 신경회로망 기반의 음성의 화자 분류 알고리즘을 제안한다. 제안한 알고리즘을 기초로 하여 연속음성 데이터베이스를 사용하여 깨끗한 음성에 배경잡음을 혼합한 후 화자 인식을 실시한다.

Ⅱ. 본 론

일반적으로 배경잡음으로 인하여 음성인식시스템의 성능이 떨어지는 문제점이 자주 발생된다[1, 3, 4]. 특히 이러한 원인으로는 주변의 배경잡음, 마이크의 특성 및 마이크와의 거리 문제 등의 여러 요소들이 음성인식 성능을 떨어뜨리는 요인이 된다. 따라서 본 논문에서는 이러한 문제를 해결하기 위하여 잡음이 섞인 환경에서 위너필터 방법이 적용된 음성인식 알고리즘을 제안한다. 음성의 분류는 다층 퍼셉트론 신경회로망(Multi-Perceptron Neural Network, MLP)[6]을 이용하여 음성인식 시스템을 구축하였다.

본 논문에서 사용한 일본어 음성 데이터베이스는 일본 음성정보처리 개발협회에서 배포한 연

구용 연속음성 데이터베이스 중에서 성인남성 화자와 성인여성 화자에 의한 문장을 임의적으로 선택하였다. 본 실험에서는 네트워크에 입력되는 MFCC는 14차(256 샘플, 32ms)의 켑스트럼 계수를 사용하며, 이 때의 중간층으로는 30의 유닛을 가지는 네트워크를 사용한다. 출력층으로는 3개의 화자를 식별하기 위하여 3개의 출력층 유닛으로 구성된다. 즉, 256 샘플의 14차에 대해서 14-30-3의 네트워크로 구성된다.

본 실험의 잡음에 대한 강건성을 테스트하기 위하여 배경잡음 환경에서 화자식별 실험을 실시하였다. 화자의존의 경우 백색잡음의 경우에 평균 화자식별률이 99% 이상으로 양호하였다. 따라서 본 논문에서 제안한 방법의 인식 성능이 효과적인 것을 실험으로 확인 할 수 있었다.

III. 결 론

본 논문에서는 주변의 배경잡음 때문에 음성인식 시스템의 성능이 떨어지는 문제점을 해결하기 위하여 위너필터 방법을 이용하여 적절한 MFCC 특징 파라미터를 추출하였다. 추출한 MFCC 특징 파라미터를 이용하여 다층 퍼셉트론 신경회로망을 기반으로 하여 음성인식 알고리즘을 제안하였다. 실험결과로부터 백색잡음이 혼합된 경우의 음성의 인식률이 효과적인 것을 알 수 있었다.

참고문헌

[1] T. Yamada, M. Kumakura and N. Kitawaki, "Performance Estimation of Speech Recognition System Under Noise Conditions Using Objective Quality Measures and Artificial Voice," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 6, pp. 2006-2013, October 2006.

[2] P. Day and A. K. Nandi, "Robust Text-Independent Speaker Verification Using Genetic Programming," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 1, pp. 285-295, January 2007.

[3] L. R. Gottlieb and G. Friedland, "On the Use of Artificial Conversation Data for Speaker Recognition in Cars," IEEE International Conference on Semantic Computing, pp. 124-128, Sept. 2009.

[4] J. Chen, J. Benesty, Y. Huang and S. Doclo, "New insights into the noise reduction Wiener filter," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, pp. 1218-1234, July 2006.

[5] W. W. Hung and H. C. Wang, "On the use of weighted filter bank analysis for the derivation of robust MFCCs," IEEE Signal Processing Letters, Vol. 8, No. 3, pp. 70-73, March 2001.

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