
음성인식 시스템의 성능 향상을 위한 잡음음성의 남성 및 여성화자의 음성식별

최재승*

*신라대학교

Speech Identification of Male and Female Speakers in Noisy Speech for Improving Performance of Speech Recognition System

Jae-seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 음성인식 알고리즘에 매우 중요한 정보를 제공하는 화자의 성별인식을 위하여 신경 회로망을 사용하여 잡음 환경 하에서 남성음성 및 여성음성의 화자를 식별하는 성별인식 알고리즘을 제안한다. 본 논문에서 제안하는 신경회로망은 MFCC의 계수를 사용하여 음성의 각 구간에서 남성음성 및 여성음성의 화자를 인식할 수 있는 알고리즘이다. 실험결과로부터 백색잡음이 중첩된 잡음환경 하에서 음성신호의 MFCC의 특징벡터를 사용함으로써 남성음성 및 여성음성의 화자에 대해서 양호한 성별인식 결과가 구해졌다.

키워드

화자인식 기법, 멜 주파수 켈프스트럼 계수, 백색잡음, 신경회로망

1. 서 론

최근 여러 연구자들에 의하여 음성인식 알고리즘에 중요한 정보를 제공하고 있는 화자의 성별인식 알고리즘이 여러 방면의 연구자들에 의해서 제안되고 있다[1]. 이러한 남성음성 및 여성음성에 대한 화자의 성별인식을 위한 식별 방법으로 가우시안 혼합 모델, 딥 신경회로망, 다층 퍼셉트론 신경회로망 등이 제안되어 있다[2, 3, 4]. 이러한 성별인식을 위한 특징 벡터로는 선형예측계수(Linear Prediction Coefficient, LPC), 멜 주파수 켈프스트럼(Mel-frequency Cepstral Coefficient,

MFCC) 등의 특징 벡터가 사용되고 있다[5, 6].

최근에 여러 잡음환경 하에서 음성인식에 대한 연구들이 다양한 방면에서 이루어지고 있으며, 특히 음성신호의 강조, 배경잡음의 제거 등 여러 연구가 이루어져 왔다. 본 논문에서는 남성음성 및 여성음성의 화자의 성별을 식별하기 위하여 다층 퍼셉트론 신경회로망(Multilayer Perceptron Neural Network, MLPNN)[7] 및 MFCC를 사용하여 백색잡음으로 오염된 음성 중에서 화자의 인식률을 향상시키는 식별 알고리즘을 제안한다. 본 논문에서 제안하는 다층 퍼셉트론 신경회로망은 멜 주파수 켈프스트럼 계수의 음성의 특징벡터를 사용

하여 음성의 각 구간에서 남성음성 및 여성음성의 화자를 분류하여 식별할 수 있는 알고리즘이다.

II. 제안한 식별 알고리즘

본 논문에서 제안한 화자의 성별식별 알고리즘은 남성음성 및 여성음성의 화자를 식별하기 위한 알고리즘이다. 본 논문에서 제안한 알고리즘은 백색잡음 하의 음성 중에서 MFCC 계수의 음성의 특징벡터를 추출한다. 이후에 MFCC 계수를 사용하여 학습 및 테스트를 실시한다.

본 실험에서는 다층 퍼셉트론 신경회로망에 입력되는 MFCC 계수는 12차이며, 제안한 12-20-2 네트워크와 12-30-2의 네트워크를 사용하여 학습시킨다. 본 논문의 실험에서 백색잡음에 대한 MFCC 계수의 평균성별 인식률은 97.16%로 양호하였다. 따라서 제안한 멜 주파수 켈스트럼 계수 및 다층 퍼셉트론 신경회로망을 사용한 경우에 화자의 평균 성별인식률이 양호한 것을 알 수 있었다.

III. 결 론

본 논문에서는 MFCC 계수 및 다층 퍼셉트론 신경회로망을 사용하여 남성음성 및 여성음성의 화자 성별식별 알고리즘을 제안하였다. 본 실험에서는 백색잡음이 중첩된 남성 및 여성음성의 화자에 대하여 다층 퍼셉트론 신경회로망을 사용하여 실험을 수행하였다. 실험 결과로부터 백색잡음에 대해서는 양호한 식별율을 구할 수 있었다. 따라서 본 논문에서 제안한 알고리즘이 백색잡음에 대하여 비교적 유효하다는 것을 실험으로 확인할 수 있었다.

참고문헌

[1] J. S. Choi, "Classification Algorithm of Male and Female in the Noisy Environment," The Journal of Korean Institute of Information Technology, Vol. 10, No. 11, pp. 63-68, Nov. 2012.
 [2] M. K. Omar, M. Hasegawa-Johnson and S. Levinson, "Gaussian mixture models of phonetic boundaries for

speech recognition," IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 33-36, Dec. 2001.
 [3] Z. Qawaqneh, A. A. Mallouh and B. D. Barkana, "Deep neural network framework and transformed MFCCs for speaker's age and gender classification," Knowledge-Based Systems, Vol. 115, pp. 5-14, Jan. 2017.
 [4] J. S. Choi, "Improved Gender Recognition System for Male and Female Speakers using MFCC," Journal of Korean Institute of Information Technology, Vol. 15, No. 9, pp. 23-28, Sep. 30, 2017.
 [5] Y. M. Zeng, Z. Y. Wu, T. Falk and W. Y. Chan, "Robust GMM Based Gender Classification using Pitch and RASTA-PLP Parameters of Speech," International Conference on Machine Learning and Cybernetics, pp. 3376-3379, August 2006.
 [6] K. V. Veena and M. Dominic, "Speaker Identification and Verification of Noisy Speech Using Multitaper MFCC and Gaussian Models", IEEE International Conference on Power, Instrumentation, Control and Computing, pp. 1-4, Dec. 2015.
 [7] 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.