
위너필터 방법을 사용한 음성 특징 벡터 추출에 의한 화자인식 기법

최재승*

*신라대학교

Speaker Recognition Technique by Extracting Speech Feature Vector using Wiener Filter Method

Jae-seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

음성인식의 적절한 성능을 구하기 위하여 잡음환경 하에서 최적인 음성의 특징 벡터를 선택할 필요가 있다. 본 논문에서는 위너필터 방법과 인간의 청각계의 특성을 활용한 멜 주파수 켈스트럼 계수를 사용한 음성인식 방법을 제안한다. 본 논문에서 제안하는 음성의 특징 벡터는 음성 중에서 배경잡음을 제거한 후에 깨끗한 음성신호의 벡터를 추출하는 방법이며, 다층 퍼셉트론 신경회로망에 멜 주파수 켈스트럼 계수를 입력하여 학습시킴으로써 음성인식을 구현한다. 본 실험에서는 멜 주파수 켈스트럼 계수의 특징 벡터를 사용하여 백색잡음이 혼합된 경우에 대하여 음성인식 실험을 실시하였다.

키워드

멜 주파수 켈스트럼 계수, 위너필터, 음성인식, 화자인식, 특징 파라미터

I. 서 론

근래에 음성의 개인 화자성 및 인간의 감정 등을 염두에 둔 음성인식 시스템 개발의 연구들이 진행되고 있다. 이러한 음성인식 시스템을 실현하는데 있어서 실제 환경의 주변잡음으로 인하여 음성인식 시스템의 성능이 저하되는 문제점이 발생되고 있다[1, 2].

이러한 다양한 잡음환경에서 음성인식 시스템의 성능을 향상시키기 위하여 음성의 특징 벡터 향상 기법, 음성강조 기법, 위너필터 기법 등이 제

안되고 있다[3, 4, 5, 6].

본 논문에서는 음성 중에서 잡음을 제거하기 위한 위너필터를 사용하여 멜 주파수 켈스트럼 계수(Mel-frequency Cepstral Coefficient, MFCC)[[7, 8]의 특징 벡터를 이용한 다층 퍼셉트론 신경회로망 기반의 음성의 화자 분류 알고리즘을 제안한다.

II. 본 론

본 논문에서는 잡음 중의 음성신호로부터 MFCC

특징 벡터를 추출한다. 따라서 이 MFCC 특징 벡터를 구하기 위하여, 먼저 위너필터 방법을 적용하여 백색잡음의 배경잡음을 제거한 후에 깨끗한 음성신호를 추출하는 방법을 제안한다. 이 위너필터 방법은 각 프레임의 음성신호로부터 배경잡음의 스펙트럼 추정하여 위너필터에 의하여 잡음을 제거한다.

음성에 대한 화자의 인식은 다층 퍼셉트론 신경 회로망(Multi-Perceptron Neural Network, MLP)[9]을 이용한 화자인식 알고리즘을 구축하였다. 본 논문에서는 14차의 MFCC 특징벡터를 사용하여 원래의 음성 및 백색잡음 하에서 음성신호에 대하여 화자를 인식하는 실험을 한다. 본 실험에서는 다층 퍼셉트론 신경회로망의 학습은 잡음이 중첩된 음성신호로부터 위너필터 방법에 의하여 음성이 향상된 MFCC 계수를 다층 퍼셉트론 신경 회로망에 입력하여 학습시킨다. 본 실험에서의 화자 분류는 각 프레임에서 다층 퍼셉트론 신경회로망에 의해서 분류된다.

본 실험에서는 백색잡음의 배경잡음 환경에서 화자인식 실험을 실시하였다. 본 실험에서는 기존의 Multitaper MFCC와 Gaussian Model[8]을 사용하여 본 논문에서 제안한 방법과 평균 화자인식률의 결과를 비교하였다. 실험의 성능비교를 통하여 본 논문에서 제안한 알고리즘은 기존의 방법보다도 7.81% 인식률이 향상되었다. 따라서 본 논문에서 제안한 방법의 성능이 기존의 방법보다 양호한 것을 알 수 있었다.

III. 결 론

본 논문에서는 백색잡음으로 오염된 잡음환경 하에서 음성인식 시스템의 성능이 떨어지는 문제점을 해결하기 위하여 위너필터 방법이 적용된 MFCC를 사용한 음성의 특징 벡터를 추출한 후에 화자인식을 하는 알고리즘을 제안하였다.

실험결과로부터 기존의 방법보다 본 논문에서 제안한 방법의 화자 인식률이 더 양호하다는 것을 확인할 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] 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.
- [3] H. K. Kim, S. H. Choi and H. S. Lee, "On approximating line spectral frequencies to LPC cepstral coefficients," *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 2, pp. 195-199, March 2000.
- [4] J. L. Carmona, J. Barker, A. M. Gómez and Ning Ma, "Speech Spectral Envelope Enhancement by HMM-Based Analysis/Resynthesis," *IEEE Signal Processing Letters*, Vol. 20, No. 6, pp. 563-566, June 2013.
- [5] 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.
- [6] J. S. Choi, "Speaker Independent Recognition Algorithm based on Parameter Extraction by MFCC applied Wiener Filter Method," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 21, No. 6, 1149-1154 Jun. 2017.
- [7] 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, Mar. 2001.
- [8] 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.
- [9] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Transaction on Neural Networks*, Vol. 3, No. 5, pp. 683-697, Sep. 1992.
- [1] T. Yamada, M. Kumakura and N. Kitawaki, "Performance Estimation of Speech Recognition