
음성 인식을 향상 위한 음성의 특징 파라미터 추출 알고리즘

최재승*

*신라대학교

Jae-Seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 잡음에 강인하고 음성인식 성능이 효과적인 멜 주파수 켈스트럼 계수의 파라미터의 추출 알고리즘을 제안한다. 본 논문에서 제안한 알고리즘은 배경잡음이 혼합된 깨끗한 연속음성 중에서 위너필터를 이용하여 음성에 포함된 배경잡음을 감소시키며, 이후에 멜 주파수 켈스트럼 계수의 특징추출 방법을 사용하여 음성의 특징 파라미터를 추출한다.

키워드

특징파라미터, 멜 주파수 켈스트럼, 배경잡음, 음성인식

I. 서 론

최근에 음성을 이용한 음성인식 시스템은 음성으로부터 단순히 음성의 언어정보만을 추출하지 않고 각 음성에 대한 개인 화자 정보 및 인간의 감정 등에 대한 화자의 특징을 정확히 추출하여 효율적인 대화를 수행하려고 하는 연구가 수행되고 있다[1, 2]. 또한 음성 중의 다양한 잡음에 대한 대처방안으로서 음성인식 시스템의 성능을 향상시키기 위하여 음성강조[3], 위너필터[4] 등의 연구들이 제안되고 있다.

본 논문에서는 잡음에 강인한 멜 주파수 켈스트럼 계수(Mel Frequency Cepstrum Coefficient, MFCC)[5]의 음성특징 파라미터를 추출하는 알고리즘을 제안한다.

II. 본 론

일반적으로 음성인식에 사용하는 특징 파라미터로는 푸리에 고속 변환(Fast Fourier Transform, FFT)에 의한 FFT 켈스트럼 계수, 선형예측부호화(Linear Predictive Coding, LPC) 분석을 통하여 파라미터를 추출하는 LPC 켈스트럼 계수 등과 같은 방법이 존재한다. 그러나 이러한 일반적인 음성의 특징파라미터와는 다르게 본 논문에서는 음성인식에 뛰어난 성능을 가진 MFCC 계수의 특징 파라미터를 제안한다. 이 MFCC는 음성의 스펙트럼 기반을 특징으로 하며 인간의 귀가 가지는 비선형적인 주파수 특성을 이용하기 때문에 유효한 특징 파라미터 값으로 알려져 있다[5]. MFCC 계수 추출 방법은 사람의 귀가 주파수 변화에 반응하는 모양이 선형적이 아닌 로그스케일과 비슷한 멜(Mel) 스케일을 따르는 청각적 특성을 반영한 켈스트럼 계수 추출 방법이다.

본 논문에서는 위너필터 방법을 적용한 MFCC

계수의 추출 알고리즘을 제안한다. 본 알고리즘은 배경잡음을 제거한 후에 깨끗한 음성신호를 추출하는 새로운 알고리즘이며, 각 프레임에서 음성신호로부터 잡음의 스펙트럼 추정하여 위너필터에 의하여 잡음을 제거한다. 이 위너필터에 의하여 기존의 MFCC 계수보다도 더 깨끗한 음성신호를 가진 MFCC 계수를 구할 수 있다.

본 논문에서는 14차의 MFCC 특징벡터를 사용하여 원음 및 백색잡음 하에서 연속음성에 대하여 음성의 특징 파라미터를 추출하는 실험을 하였다. 본 논문에서 사용한 일본어 음성 데이터베이스는 일본 음성정보처리 개발협회에서 배포한 연구용 연속음성 데이터베이스 중에서 성인남성 화자와 성인여성 화자에 의한 문장을 임의적으로 선택하여 특징파라미터를 추출한다.

본 실험에서는 MFCC 계수는 14개의 특징벡터를 추출하였으며, 이 추출된 특징벡터를 이용하여 향후에 음성의 화자분류 및 음성인식을 하는 기초적인 연구를 목적으로 한다. 따라서 본 논문에서는 인간의 귀의 비선형적 특성을 사용한 MFCC 특징 파라미터를 이용함으로써 정확한 특징 벡터를 추출할 수 있었다.

III. 결 론

본 논문에서는 배경잡음으로 인하여 음성인식 시스템의 성능이 크게 저하되는 문제점을 해결하기 위하여 위너필터 방법을 이용한 MFCC 특징 파라미터를 추출하였다. 이 추출된 MFCC 특징 파라미터를 이용하여 향후의 신경회로망 등을 이용하여 음성인식 및 화자분류 등의 기초연구에 활용할 수 있다.

참고문헌

- [1] 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.
- [2] P. Song, Y. Jin, C. Zha and L. Zhao, "Speech emotion recognition method based on hidden factor analysis," *Electronics Letters*,

Vol. 51, No. 1, pp. 112-114, January 2015.

- [3] 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.
- [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] 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.