

<심사용 논문>

자동차 잡음 환경에서 웨이브렛 밴드 엔트로피 앙상블 분석을 이용한 음성구간 검출 알고리즘

*이 기 현, **김 명 남

*경북대학교 대학원 의용생체공학과

**경북대학교 의학전문대학원 의공학교실

Voice Activity Detection Algorithm using Wavelet Band Entropy Ensemble Analysis in Car Noisy Environments

*Gihyun. Lee, **Myoung Nam Kim

*Dept. of Medical & Biological Eng., Graduate School, Kyungpook National
University

**Dept. of Biomedical Eng., School of Medicine, Kyungpook National
University

● 투고분야 : 신호 및 영상처리

● Corresponding Author : Myoung Nam Kim

- Address : (700-842) Gukchaebosang-ro 680, Jung-gu, Daegu, Korea
- TEL : +82-53-200-XXXX
- FAX : +82-53-200-XXXX
- E-mail : kkkkkk@knu.ac.kr

● This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-000XXXX)

※ 교신저자 및 연구비 지원사항은 모두 영문으로 작성되어야 함!!!!

자동차 잡음 환경에서 웨이브렛 밴드 엔트로피 앙상블 분석을 이용한 음성구간 검출 알고리즘

Voice Activity Detection Algorithm using Wavelet Band Entropy Ensemble Analysis in Car Noisy Environments

Abstract

Voice activity detection is very important process that voice activity separated form noisy speech signal for speech enhance. Over the past few years, many studies have been made on voice activity detection, but it has poor performance in low signal to noise ratio environment or fickle noise such as car noise. In this paper, it proposed new voice activity detection algorithm using ensemble variance based on wavelet band entropy and soft thresholding method. We conduct a survey in a lot of signal to noise ratio environment of car noise to evaluate performance of the proposed algorithm and confirmed performance of the proposed algorithm.

Keyword : Voice Activity Detection, Entropy, Wavelet Band, Ensemble, Car Noise

※ 초록과 키워드는 반드시 영문으로 작성하고 초록은 반드시 200단어이내이어야 함.
단어 수를 확인하기 바람!!!! (키워드의 각 단어의 첫글자는 대문자임)

1. 서론

최근 스마트기기에 음성인식 기술을 접목하여 음성명령이나 음성비서와 같은 여러 가지 새로운 편리한 기능이 소개되면서 여러 분야에서 음성인식에 관한 관심이 증가하고 있다. 그리고 스마트폰뿐만 아니라 Tablet-PC, 가전제품 등의 제품에서도 음성인식을 이용한 음성 인터페이스 시스템을 탑재하려는 연구가 활발히 이루어지고 있으며 특히 스마트 자동차에 대한 연구가 많은 기대를 받고 있다. 스마트 자동차의 음성 인터페이스 시스템을 이용하여 주행 중에 차량의 상태를 음성으로 명령하고 보고받거나 음성명령으로 길 안내를 받는 등의 기능을 수행할 수 있기 때문이다. 하지만 지금까지의 음성인식 기술은 조용한 환경에서 단음절 혹은 짧은 단어에 대해서는 인식률이 좋은 편이지만 잡음이 많은 환경이나 문장단위의 음성에 대해서는 인식률이 낮았다. 음성 인식률을 높이기 위해서 많은 기술들이 개발되어야 하지만 그 중 먼저 선행되어야 하는 중요한 기술은 음성구간 검출 기법이다.

음성구간 검출 기법은 에너지 값, 영교차율(zero crossing rate), 선형 예측 부호화(linear predictive coding, LPC) 계수(parameter) 등과 같은 양적 특징들을 이용하는 방법과 우도비

(likelihood ratio, LR), 엔트로피(entropy) 등과 같은 통계적 특징들을 이용하는 방법이 있다[1]. 그 중 신호의 에너지를 이용한 방법은 신호 대 잡음비(signal to noise ratio, SNR)이 낮은 환경에서 성능이 급격히 저하되는 단점이 있으며 영교차율은 특정잡음에 대해서는 음성과 잡음을 구분하지 못하는 단점이 있었다[2]. 그리고 LR와 같은 통계적인 특징을 이용하는 연구들의 경우, 좋은 성능을 보여주고 있지만 계산량이 많거나 음성과 통계적 특징이 비슷한 잡음에서는 좋은 성능을 보여주지 못하였다[3-6]. 최근 다양한 잡음환경에서 사용이 가능하고 통계적 특성을 이용하는 엔트로피를 이용한 음성검출 방법에 대한 연구가 활발히 이루어지고 있다[2]. Asgari[7]는 주파수 영역에서의 엔트로피를 이용한 음성구간 검출 방법을 제안하였다. 다양한 SNR 환경에 대해 실험하였는데 높은 SNR환경의 대해서는 좋은 성능을 보였으나 SNR이 낮아지면 성능이 저하되는 단점을 보였다.

본 논문에서는 ‘시간-스케일’과 ‘주파수-스케일’영역을 함께 분석하기에 적절한 웨이브렛 변환(wavelet transform)을 사용하여 밴드(band)를 나누어 엔트로피를 구한 다음, 웨이브렛 밴드 엔트로피에 기반한 앙상블 분산(ensemble variance)을 음성구간 검출 신호로 사용하여 낮은 SNR에서도 음성구간을 효과적으로 강조시켰다. 그리고 다양한 SNR과 잡음의 변화에 적응시켜 음성구간을 검출하기 위해 기존의 하드 문턱치(hard thresholding) 기법 대신에 엔트로피에 기반한 새로운 소프트 문턱치(soft thresholding) 기법을 제안하였다. 제안한 알고리즘은 유효성을 평가하기 위하여 기존의 알고리즘이 좋은 성능을 보여주지 못한 자동차 잡음을 대상으로 실험을 수행하였다. 이는 자동차 잡음이 음성과 비슷한 특성을 가지며 시간에 따른 변화가 크기 때문이다. 그리고 실제 조용한 자동차 안이나 시끄러운 도로변의 상황과 같은 다양한 상황을 감안하여 실험하기 위하여 다양한 SNR환경에서 실험을 수행하였다. 실험 결과, 제안한 방법이 기존의 방법에 비하여 성능이 우수함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문의 기본 이론이 되는 웨이브렛 변환과 엔트로피에 대하여 정의한 다음, 3장에서 제안하는 음성구간 검출 알고리즘을 설명한다. 여기에서는 제안한 알고리즘의 전체적인 개요와 웨이브렛 분해, 웨이브렛 밴드 엔트로피, 웨이브렛 밴드 엔트로피 앙상블 분산, 그리고 소프트 문턱치 음성구간 검출법의 순서로 서술한다. 그리고 4장에서는 제안한 알고리즘으로 실험한 결과를 기존의 알고리즘들과 객관적인 지표로 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

※ 참고문헌 표기 순서를 지킬 것 ([1]부터 차례로 순차적으로 부여)

2. 이론

2.1 웨이브렛 변환

웨이브렛 해석은 웨이브렛이라고 불리는 하나의 원형 함수와 이 함수의 스케일(scale)된 함수가 기저함수(basis function)를 이루게 되며 대역통과필터와 같은 역할을 수행하게 된다[8]. 웨이브렛 해석에는 다양한 웨이브렛 함수가 활용될 수 있으며 일반적으로 대상 신호의 특성에 맞는 함수를 선택하여 적용한다. 음성신호처리 분야에서는 음성과형과 유사한 모양을 가지는 daubechies 웨이브렛 함수를 많이 사용한다[9]. 신호 x 의 웨이브렛 변환은 다음과 같이 정의된다.

$$\Psi_x(a,b) = \frac{1}{a} \int_{-\infty}^{+\infty} h^*\left(\frac{t-b}{a}\right) x(t) dt \quad (1)$$

※ 모든 수식은 영문과 수학기호로 작성되어야하며 수식번호도 부여되어야 함!!!!

여기서 a 는 스케일 계수(scale parameter)이며 b 는 이동 계수(shift parameter)이다. 식 (1)에서 $h_{a,b}(t)$ 는 웨이브렛 모함수(mother wavelet function) $h(t)$ 로부터 a 값에 따라 확장 및 수축하게 되며 다음과 같이 정의된다.

2.2 엔트로피

엔트로피는 열역학적으로 ‘통계적인 무질서도’를 의미하며 정보통신이나 신호처리 분야에서는 ‘데이터에 내재되어 있는 정보의 양’을 나타낸다[9]. 음성신호 데이터에 내재되어 있는 정보의 양을 시간에 따라 나타낼 수 있으므로 시간영역에서 음성신호의 엔트로피를 정의 할 수 있다.

3. 제안한 방법

3.1 제안한 알고리즘의 개요

본 논문에서는 자동차 잡음이 섞인 음성신호에서 음성구간을 검출하기 위한 음성구간 검출 알고리즘을 제안한다. Fig. 1에 제안하는 음성구간 검출 알고리즘의 전체적인 흐름도를 나타내었다.

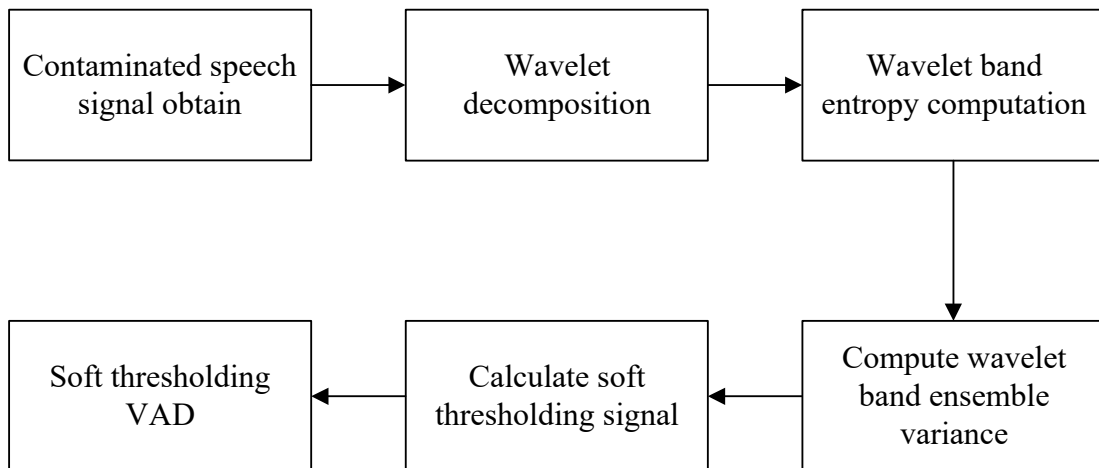


Fig. 1. Flowchart for voice activity detection algorithm.

※ 그림의 제목과 내용은 모두 영문으로 작성!!!! 한 줄일 경우 가운데 정렬, 두 줄 이상일 경우 왼쪽 정렬로 작성

먼저, 잡음이 섞인 음성신호가 입력이 되면 이 음성신호를 여러 개의 웨이브렛 밴드(band)로 나누는 웨이브렛 분해과정을 거친다. 그리고 각각의 웨이브렛 밴드의 엔트로피를 계산한 후, 웨이브렛 밴드에 대한 양상블 분석과정을 거친다. 양상블 분석과정을 통해 얻은 양상블 특징신호들을 합성하면 음성구간 검출을 위한 양상블 합성신호를 얻게 된다. 이상을 과정을 통해 얻은 양상블 합성신호를 이용해 소프트 문턱치 검출법으로 주어진 음성신호의 음성구간을 검출한다.

4. 실험 결과 및 고찰

4.1 음성구간 검출 결과

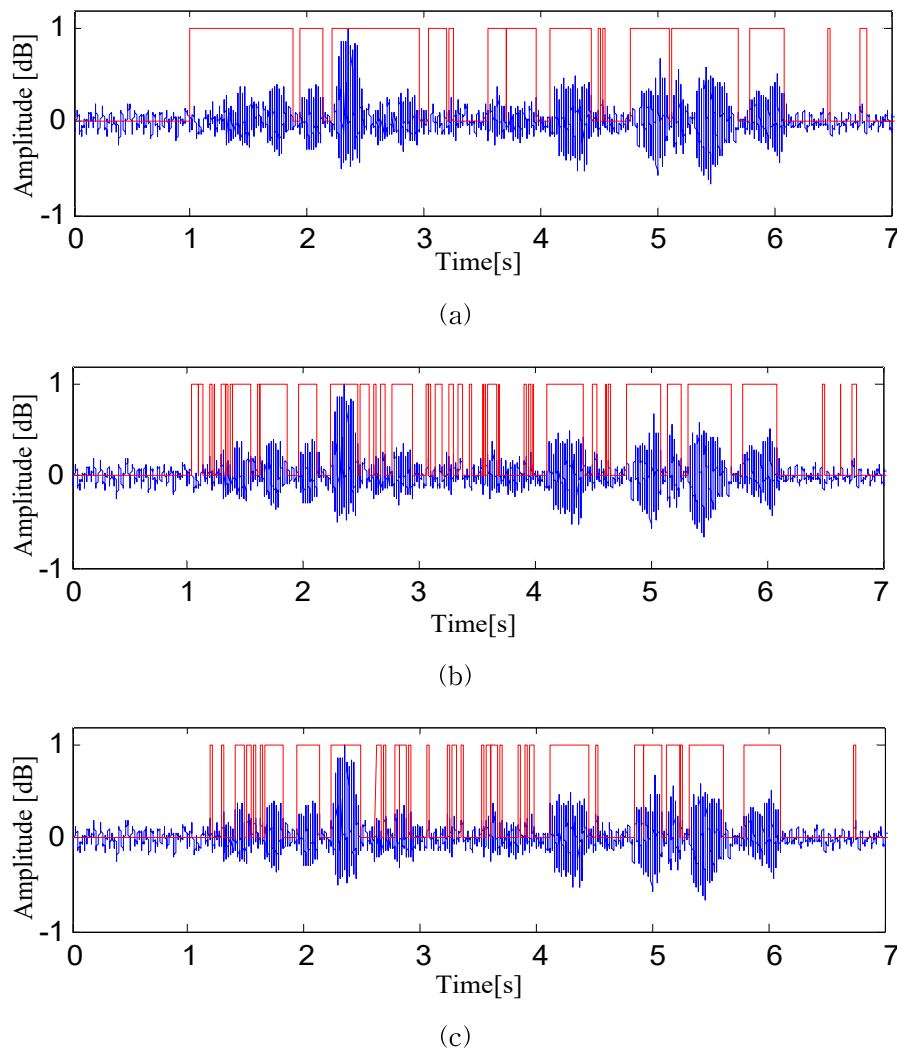


Fig. 7. Detected voice regions using conventional algorithm and proposed algorithm in case of SNR 0dB. (a) MAE, (b) EVAD, and (c) proposed algorithm.

※ 모든 각각의 그림에는 (a), (b).... 등 코드를 주고 제목에서 (a), (b)의 설명이 제시

되어야 함!!!!

※ 모든 그래프에는 가로축과 세로축에 대한 명칭과 단위가 표시되어야 함.

Table 1. The performance for proposed algorithm and conventional algorithms.

SNR (dB)	Algorithm		
	MAE	EVAD	Proposed
	PHR[%]	PHR[%]	PHR[%]
0	56.52	60.83	78.88
5	56.01	60.79	77.62
10	55.24	60.89	79.62
15	56.87	63.49	80.85

※ 표의 제목과 내용은 모두 영문으로 작성!!!!

5. 결 론

본 논문에서는 새로운 웨이브렛 밴드 엔트로피 기반의 양상블 분산과 소프트 문턱치를 이용한 음성검출 알고리즘을 제안하였다.

REFERENCE

※ 모든 참고문헌은 영문으로 작성!! 국문 참고문헌일지라도 반드시 영문으로 작성한다.

※ 저자명 표기법은 이름 다음에 성이 오도록 하며 이름은 첫 문자의 약어로 표현하며 (예를 들면, Shi-Wen(이름) Deng(성) 일 경우 S. Deng으로 Shi(첫번째 이름) Wen(두번째 이름) Deng(성) 일 경우 S.W. Deng로 표기), 저자가 6명이상일 경우 6 명의 성명만 나열하고 그 다음에 “et al.” 로 표기한다.

예시) Author1, Author2,, Author6, et al., “Title of the Paper,” Journal...

※ 논문 저자가 2명일 경우 으로 쉼표(,)없이 표기. 3명 이상일 경우 D.A. Adjero, M. Ryynanen, and K.C. Nwosu 으로 and 앞에 쉼표(,) 표기

예시) 2명일 경우 : D.A. Adjero and K.C. Nwosu, “Title,” Journal, ...

3명 이상일 경우 :D.A. Adjero, M. Ryynanen, and K.C. Nwosu, “Title,”

Journal, ...

- ※ 논문제목, 논문지명은 전치사 및 접속사 등을 제외하고는 모든 단어의 첫글자는 대문자로 표기하며 논문지명, 도서명, 보고서명, 학술회의 초록명 등은 이탤릭체로 표기한다.
- ※ 논문지명, 학술회의명은 약어를 사용하지 않고 전체를 풀어서 작성한다.
- ※ Open access online 논문은 탐색 가능한 수준으로 최대한의 정보 기입한다.
- ※ 기타 설명되지 않은 부분은 아래의 형식을 준용한다.

- [1] D.A. Adjero, M. Ryynanen, and K.C. Nwosu, "Multimedia Database Management Issues," *Journal of Korea Multimedia Society*, Vol. 4, No. 3, pp. 24-33, 1997. (논문지의 경우)
- [2] S. Khoshafian, K.C. Nwosu, and B.A. Baker, *Multimedia and Imaging Databases*, Morgan Kaufmann Publishers, San Francisco, Calif., 1996. (단행본의 경우)
- [3] M. Ryynanen, D.A. Adjero, and A. Klapuri, "Query by Humming of MIDI and Audio using Locality Sensitive Hashing," *Proceeding of International Conference Acoustic, Speech and Signal Processing*, pp. 2249-2252, 2008. (학술회의 초록의 경우)
- [4] M. Ryynanen and A. Klapuri, "Query by Humming of MIDI and Audio using Locality Sensitive Hashing," *Proceeding of the Fall Conference of the Korea Multimedia Society*, pp. 2249-2252, 2008. (국내학술대회 초록의 경우)
- [5] T. Makela, *Dental X-ray Image Stitching Algorithm*, Master's Thesis of Helsinki University of Technology, 2009. (학위논문의 경우)
- [6] The Story on the PING Program(2008), <http://ftp.arl.mil/~mike/ping/html> (accessed January 24, 2012). (인터넷의 경우)
- [7] ETSI, *Radio Broadcasting Systems: Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers*, ETSI EN 300 401, v.1.4.1, 2006. (보고서 등의 경우)
- [8] LG Electronics, *Display Device and Channel Composition Method*, 10-0724644, Korea, 2007. (특허의 경우)
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv Preprint*, arXiv:1512.03385, 2015. (Open Access Online 논문 중 arXiv 논문일 경우)
- [10] H.-K. Lam, P.-D. Lam, and S.-H. Lee, "Digital Twin Smart City Visualization with MoE-Based Personal Thermal Comfort Analysis," *Sensors*, Vol. 25, No. 3, Article 705, 2025. (페이지 번호 대신 Article 번호 사용하는 논문일 경우)
- [11] S. Ansolabehere, M. Palmer, and A. Lee. "Precinct-Level Election Data," 2014, Distributed by Harvard Election Data Archive, <http://hdl.handle.net/1902.1/21919> UNF:5:5C9UfGjdLy2ONVPtgr45qA= (accessed August 1, 2023). (Dataset 경우)

※ 참고문헌 표기법을 꼭 준수하여 주시길 바랍니다.