Journal of Korea Multimedia Society Vol. 16, No. 9, September 2013(pp. 1005-1017)
http://dx.doi.org/10.9717/kmms.2013.16.9.1005

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

이기현[†], 이윤정[††], 김명남[†††]

요 약

음성구간 검출은 음성과 잡음이 섞인 신호에서 음성구간과 비음성구간을 구분하는 과정으로 음성 향상을 위한 신호처리에서 매우 중요한 과정이다. 지금까지 음성구간 검출에 관한 많은 연구가 있었지만, 낮은 신호 대 잡음비 환경이나 자동차 잡음과 같은 시간에 따른 변화가 심한 잡음환경에서는 좋은 성능을 보이지 못하였다. 본 논문에서는 웨이브렛 밴드 엔트로피 기반의 앙상블 분산과 소프트 문턱치 기법을 이용한 새로운 음성구간 검출 알고리듬을 제안하였다. 제안한 알고리듬의 성능을 비교 평가하기 위하여 자동차 잡음이 있는 다양한 신호 대 잡음비 환경에서 실험을 수행하였으며 실험결과, 제안한 방법의 우수한 성능을 확인할 수 있었다.

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

G.H. Lee[†], Y.J. Lee[††], M.N. Kim[†††]

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.

Key words: Voice Activity Detection(음성구간 검출), Entropy(엔트로피), Wavelet Band(웨이브렛 밴드), Ensemble(앙상블), Car Noise(자동차 잡음)

1. 서 론

최근 스마트기기에 음성인식 기술을 접목하여 음성명령이나 음성비서와 같은 여러 가지 새롭고 편리한 기능이 소개되면서 여러 분야에서 음성인식에 관한 관심이 증가하고 있다. 그리고 스마트폰뿐만 아니라 Tablet-PC, 가전제품 등의 제품에서도 음성인식을 이용한 음성 인터페이스 시스템을 탑재하려는 연

※ 교신저자(Corresponding Author) : 김명남, 주소 : 대구광역시 중구 국채보상로 680, 경북대학교 의학전문대학원 의공학교실(700-842), 전화 : 053) 200-5266, FAX : 053) 200-5264 , E-mail : kimmn@knu.ac.kr
접수일 : 2013년 6월 19일, 수정일 : 2013년 7월 22일
완료일 : 2013년 8월 12일

[†] 경북대학교 대학원 의용생체공학과
(E-mail: gihyounlee@gmail.com)
[††] 경북대학교 대학원 의용생체공학과
(E-mail: whitegleam@nate.com)
[†††] 경북대학교 의학전문대학원 의공학교실
※ 본 연구는 보건복지가족부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임. (과제고유번호: A092106)

구가 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 음성 명령으로 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 문장 단위의 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)가 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 알고리듬의 유효성을 평가하기 위하여 기존의 알고리듬이 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. 실험결과, 제안한 방법이 기존의 방법에 비하여 성능이 우수함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 음성구간 검출 기법인 주파수 영역에서의 엔트로피를 이용한 음성구간 검출 방법에 대하여 설명한 다음, 3장에서 제안하는 음성구간 검출 알고리듬을 설명한다. 여기에서는 제안한 알고리듬의 전체적인 개요와 웨이브렛 분해, 웨이브렛 밴드 엔트로피, 웨이브렛 밴드 엔트로피 앙상블 분산, 그리고 소프트 문턱치 음성구간 검출법의 순서로 서술한다. 그리고 4장에서는 제안한 알고리듬으로 실험한 결과를 기존의 알고리듬들과 객관적인 지표로 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

2. 이론

2.1 주파수 영역에서의 엔트로피를 이용한 음성구간 검출[7]

엔트로피는 열역학적으로 '통계적인 무질서도'를 의미하며 정보통신이나 신호처리 분야에서는 '데이터에 내재되어 있는 정보의 양'을 나타낸다[8]. 음성신호 데이터에 내재되어 있는 정보의 양을 시간에 따라 나타낼 수 있으므로 시간영역에서 음성신호의 엔트로피를 정의할 수 있다. 주어진 음성신호의 시간영역에서의 특정 프레임(k)에서 엔트로피는 식 (1)과 같이 정의된다[8].

$$H_t(k) = -\sum_{i=k-l/2}^{k+l/2} p[x(i)]\log(p[x(i)]) \quad (1)$$

여기서 l은 엔트로피 계산을 위한 프레임의 길이이다. 식 (1)을 응용하여 Asgari[7]는 주파수 영역에서 엔트로피를 식 (2)와 같이 정의 하였다[7].

$$H_s(|Y(\omega,t)|) = -\sum_{\omega}^{\Omega} p[|Y(\omega,t)|]\log(p[|Y(\omega,t)|]) \quad (2)$$

여기서 $p[|Y(\omega,t)|]$는 주파수 영역에서의 특정 프레임(t)에서 특정 주파수 밴드(ω)의 확률이다. 그리고 $H(|Y(t)|)$는 백색잡음에서 최대값을 가지며 순수 주기신호에서 최소값을 가진다[7]. 각각의 프레임에 대해 고속 푸리어 변환 계수(fast fourier transform coefficient)의 히스토그램(histogram)을 이용해 주파수 영역에서의 전력스펙트럼밀도(power spectrum density, PDF)를 구한 후 식 (2)를 이용해 각 프레임의 엔트로피를 계산한다[7]. 또한 스무딩필터(smoothing filter)를 활용하여 음성 구간과 잡음 구간의 격차를 크게 만들어 문턱치 결정률을 향상 시켰다. 마지막으로 처음 30프레임을 순수한 잡음구간으로 설정하여 순수한 잡음의 PDF와 음성과 잡음이 섞인 신호의 PDF를 계산하여 두 값의 우도비 검사법(likelihood ratio test method)[9]을 활용해 최적의 문턱치를 결정하여 음성구간과 잡음구간을 최종적으로 결정하였다[7].

3. 제안하는 음성구간 검출 알고리듬

3.1 제안한 알고리듬의 개요

본 논문에서는 자동차 잡음이 섞인 음성신호에서 음성구간을 검출하기 위한 음성구간 검출 알고리듬을 제안한다. 그림 1에 제안하는 음성구간 검출 알고리듬의 전체적인 흐름도를 나타내었다.

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 앙상블 분석과정을 거친다. 앙상블 분석과정을 통해 얻은 앙상블 특징신호들을 합성하면 음성구간 검출을 위한 앙상블 합성신호를 얻게 된다. 이상을 과정을 통해 얻은 앙상블 합성신호를 이용해 소프트 문턱치 검출법으로 주어진 음성신호의 음성구간을 검출한다.

3.2 웨이브렛 분해

웨이브렛 해석은 웨이브렛이라고 불리는 하나의 원형 함수와 이 함수의 스케일(scale)된 함수가 기저함수(basis function)를 이루게 되며 대역통과필터와 같은 역할을 수행하게 된다[10]. 웨이브렛 해석에는 다양한 웨이브렛 함수가 활용될 수 있으며 일반적으로 대상 신호의 특성에 맞는 함수를 선택하여 적용한다. 음성신호처리 분야에서는 음성파형과 유사한 모양을 가지는 Daubechies 웨이브렛 함수를 많이 사용한다[11]. 신호 x의 웨이브렛 변환은 다음과 같이 정의된다.

$$\Psi_x(a,b) = \frac{1}{a}\int_{-\infty}^{+\infty} h^*\left(\frac{t-b}{a}\right)x(t)dt \tag{3}$$

여기서 a는 스케일 계수(scale parameter)이며 b는 이동 계수(shift parameter)이다. 식 (3)에서 h^*는 h의 공액복소수이며, $h_{a,b}(t)$는 웨이브렛 모함수(mother wavelet function) $h(t)$로부터 a값에 따라 확장 및 수축하게 되며 다음과 같이 정의된다.

$$h_{a,b}(t) = \frac{1}{a}h\left(\frac{t-b}{a}\right) \tag{4}$$

웨이브렛 변환이 기존의 주파수 변환 방법인 푸리에 변환으로는 표현할 수 없는 '시간-스케일' 공간을 표현할 수 있는 이유는 웨이브렛의 스케일 특성에 의한 것이며 식 (4)에서 a>1 이면 웨이브렛은 팽창하

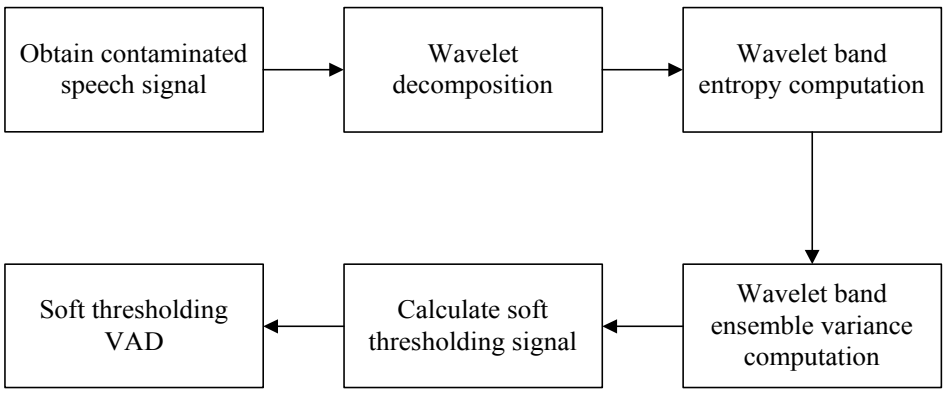

그림 1, 제안한 음성구간 검출 알고리듬 흐름도

고 a<1 이면 웨이브렛은 압축하여 '시간-스케일' 공간에 신호의 스케일 성분을 표현한다[12].

주어진 신호를 여러 개의 웨이브렛 밴드로 분해하기 위해 식 (3)을 이용해 웨이브렛 변환을 하였다. 여기서 웨이브렛 변환을 이용해 신호를 분해하는 과정은 그림 2와 같은 트리(tree) 형태로 구성된 필터뱅크(filter bank) 구조로 되어있다. 그림 2의 필터뱅크 구조는 Mallat[13]가 제안한 웨이브렛 분해 트리구조로서 H_0는 저역통과 필터이고 H_1은 고역통과 필터이다.

입력된 음성 신호가 저역통과 필터(low pass filter)와 고역통과 필터(high pass filter)를 차례로 거치면서 신호가 분해된다. 음성신호에서 급격히 변화하는 부분의 위치를 각 스케일에서 보존하기 위해 저역통과 필터를 통과한 신호는 반복적으로 웨이브렛 변환과정을 거치면서 필터링된 신호의 크기를 유지 한다[14]. 웨이브렛 분해를 위해 daubechies6 (db6) 웨이브렛 함수를 이용한 7단계 트리를 이용하였고 그림 2의 필터 뱅크 구조를 활용하여 8개의 웨이브렛 밴드로 분해하였다. 식 (3)을 이용하여 자동차 잡음에 오염된 음성신호 $y(t)$의 시간(t)에 따른 웨이브렛 변환과 8개의 웨이브렛 밴드로 분해된 신호를 다음과 같이 표현할 수 있다.

$$\Psi_y(t) = \frac{1}{a}\int_{-\infty}^{+\infty} h^*\left(\frac{t-b}{a}\right) y(t)dt \tag{5}$$

$$\psi_y^j(t) = \begin{bmatrix} \psi_y^1(t) \\ \psi_y^2(t) \\ \vdots \\ \psi_y^j(t) \end{bmatrix} = \begin{bmatrix} \psi_y^1(1)\,\psi_y^1(2)\cdots\psi_y^1(t) \\ \psi_y^2(1) \quad \ddots \quad \vdots \\ \vdots \quad \ddots \quad \vdots \\ \psi_y^j(1)\,\psi_y^j(2)\cdots\psi_y^j(t) \end{bmatrix} \tag{6}$$

여기서 $t=1,2,...,L$이고 L은 신호의 길이이다. 그리고 식 (5)의 $\Psi_y(t)$는 오염된 음성신호 $y(t)$의 시간(t)에 따른 웨이브렛 변환이며 식 (6)은 $y(t)$를 시간(t)에서 j개의 웨이브렛 밴드로 분해한 음성신호 $\psi_y^j(t)$를 행렬의 형태로 나타낸 것이며 $j=1,2,...,8$이다. 그림 3에 깨끗한 음성신호와 자동차 잡음이 섞인 음성신호를 나타내었다. 여기서 신호는 최대값이 1이 되도록 정규화 과정을 거쳤으므로 y축은 정규화된 신호의 크기를 나타내며 x축은 시간을 나타낸다. 그리고 그 자동차 잡음이 섞인 음성신호를 식 (6)을 이용하여 8개의 웨이브렛 밴드로 분해된 신호를 그림 4에 나타내었다.

그림 4에서 8개의 웨이브렛 밴드로 분해된 신호를

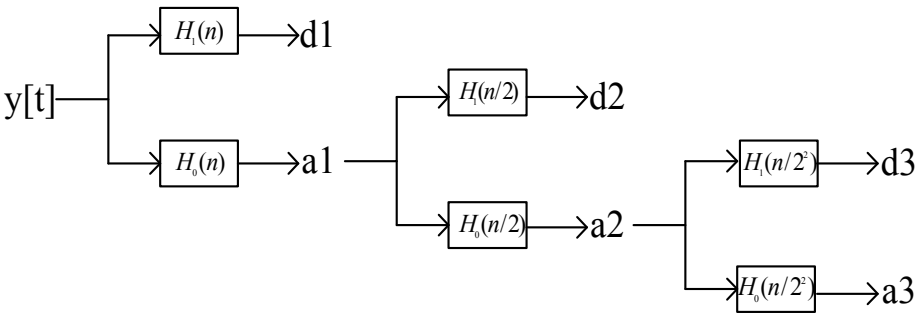

그림 2. 웨이브렛 변환에 대한 필터뱅크 구조

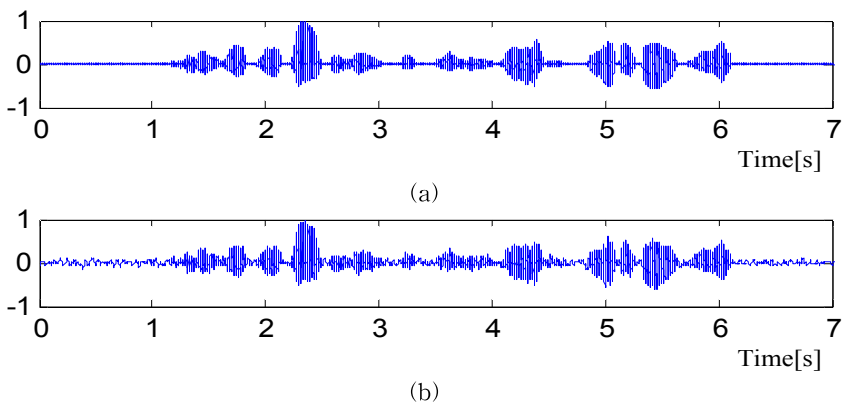

그림 3. 실험에 사용된 음성신호 (a) 순수한 음성신호와 (b) 자동차 잡음과 음성이 섞인 신호

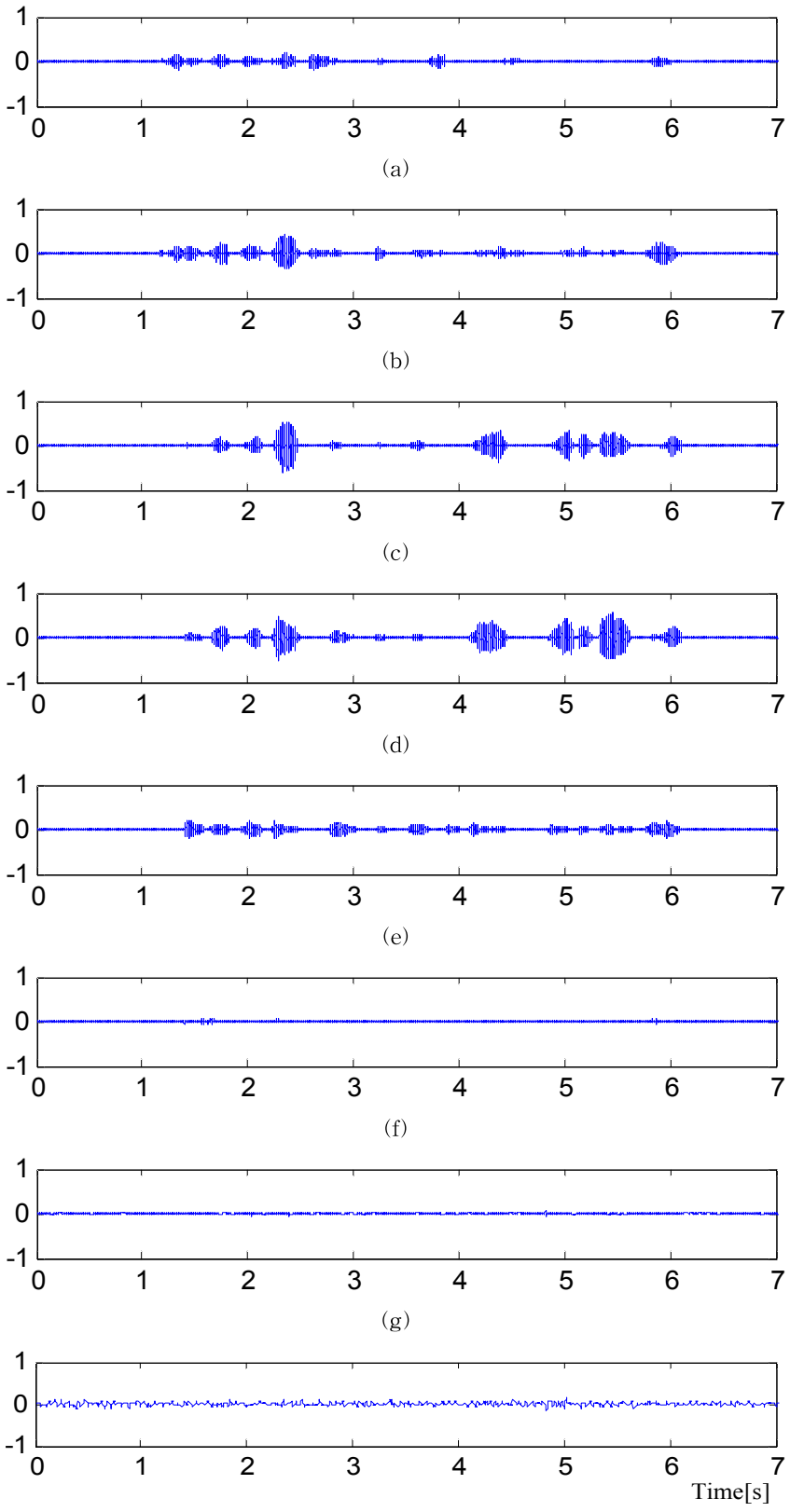

그림 4. 자동차 잡음과 음성이 섞인 신호에 대한 웨이브렛 분해방법에 의해 분해된 신호 (a) d1, (b) d2, (c) d3, (d) d4, (e) d5, (f) d6, (g) d7, (h) a7

볼 수 있다. 음성구간에서 일부 웨이브렛 밴드에서는 큰 에너지를 보이고 일부 밴드에서는 잡음과 차이가 없는 작은 에너지를 보이는 것을 볼 수 있다. 이러한 특징을 웨이브렛 밴드 엔트로피(wavelet band entropy)를 계산하여 강조시킬 수 있다.

3.3 웨이브렛 밴드 엔트로피

음성구간을 강조하고 잡음구간을 감쇄시키기 위해 웨이브렛 분해로 나누어진 신호 각각의 웨이브렛 밴드별로 엔트로피를 계산한다. 불규칙한 데이터의 정보의 양을 높게 표현하기 위해 식 (1)를 변형하여 식 (7)과 같이 나타내었다.

$$H_x(k)=\sum_{i=k-l/2}^{k+l/2} p[x(i)]\log(p[x(i)]) \tag{7}$$

그리고 식 (7)을 이용해 연속된 데이터에 대해 시간에 따라 신호의 엔트로피를 계산하기 위해 $y(t)$를 대입하여 다음과 같이 전개하였다[15].

$$\rho_{\psi_y^j}(t)=\frac{\sum_{i=t-N/2}^{t+N/2} p[\psi_y^j(i)]\log(p[\psi_y^j(i)]) - \frac{1}{L}\sum_{j=1+N/2}^{L-N/2}\left[\sum_{i=t-N/2}^{t+N/2} p[\psi_y^j(j+i)]\log(p[\psi_y^j(j+i)])\right]}{\sqrt{\frac{1}{L}\sum_{m=1}^{L}\left[\sum_{i=t-N/2}^{t+N/2} p[\psi_y^j(i)]\log(p[\psi_y^j(i)]) - \frac{1}{L}\sum_{j=1+N/2}^{L-N/2}\left[\sum_{i=t-N/2}^{t+N/2} p[\psi_y^j(j+i)]\log(p[\psi_y^j(j+i)])\right]\right]^2}} \tag{8}$$

$$\rho_{\psi_y^j}(t)=\begin{bmatrix}\rho_{\psi_y^1}(t)\\ \rho_{\psi_y^2}(t)\\ \vdots\\ \rho_{\psi_y^j}(t)\end{bmatrix}=\begin{bmatrix}\rho_{\psi_y^1}(1)\,\rho_{\psi_y^1}(2)\ldots\rho_{\psi_y^1}(t)\\ \rho_{\psi_y^2}(1)\quad\ddots\quad\vdots\\ \vdots\qquad\quad\ddots\quad\vdots\\ \rho_{\psi_y^j}(1)\,\rho_{\psi_y^j}(2)\ldots\rho_{\psi_y^j}(t)\end{bmatrix} \tag{9}$$

여기서 N은 $50ms$내의 음성신호 샘플 개수이며 식 (8)과 식 (9)의 $\rho_{\psi_y^j}(t)$는 웨이브렛 밴드로 분해한 음성신호 $\psi_y^j(t)$의 j번째 웨이브렛 밴드의 시간에 따른 엔트로피 신호의 값이다. 그리고 식 (9)는 식 (8)을 이용해 계산된 엔트로피 신호의 값을 행렬형태로 나타낸 것이다. 식 (8)과 식 (9)을 통해 계산된 웨이브렛 밴드 엔트로피 신호를 그림 5에 나타내었다.

식 (7)에서 불규칙한 구간의 엔트로피 신호의 값이 크게 나타나도록 변형하였으므로 그림 5에서 음성구간보다 비교적 불규칙한 특성을 가지는 잡음구간에서 엔트로피 신호의 값이 크게 나타난다. 웨이브렛 밴드 엔트로피 신호는 각각의 엔트로피 밴드에서 신호 데이터의 불규칙한 정보의 양을 동일한 시간대에서 볼 수 있다. 또한 음성이 많이 포함된 웨이브렛 밴드에서 음성구간의 값이 작게 나타나고 음성이 비교적 적게 포함된 웨이브렛 밴드에서는 음성구간에서도 잡음구간과 비슷한 엔트로피 신호의 값이 나타난다.

3.4 웨이브렛 밴드 엔트로피 앙상블 분산

웨이브렛 엔트로피 신호는 불규칙한 특성을 가진 잡음구간에서 큰 값을 가지고 음성구간에서 상대적으로 작은 값을 가진다. 그리고 앞서 분해한 웨이브렛 밴드를 이용해 각각의 웨이브렛 밴드의 엔트로피 신호를 계산하여 행렬의 형태로 나타내면 같은 구간에서 음성이 포함된 밴드와 음성이 포함되지 않은 밴드의 특성을 함께 볼 수 있다. 이러한 웨이브렛 밴드 엔트로피의 특성을 이용하여 같은 시간대에서 여러 개의 웨이브렛 밴드의 엔트로피를 앙상블 분석을 한다. 식 (9)의 웨이브렛 밴드 엔트로피의 행렬 형태에서 다음에 따라 앙상블 분석 특징신호를 추출한다.

$$\sigma_\epsilon(t)=\frac{1}{j}\left\{\rho_{\psi_y^n}(t)-\frac{1}{j}\sum_{n=1}^{j}\rho_{\psi_y^n}(t)\right\}^2 \tag{10}$$

여기서 j는 웨이브렛 밴드 개수를 나타낸다. 식 (10)을 통해 얻은 $\sigma_\epsilon(t)$는 시간에 따른 웨이브렛 밴드 엔트로피 앙상블 분산(wavelet band ensemble variance)이라 한다. 앙상블 분석을 통해 얻은 신호를 그림 6에 나타내었다.

그림 6(b)는 웨이브렛 밴드 엔트로피 앙상블 분산으로 웨이브렛 밴드간의 엔트로피 차이가 큰 음성구간에서 매우 큰 값을 나타내고 잡음구간에서는 상대적으로 작은 값을 나타낸다. 하지만 잡음구간 중에서 음성과 비슷한 특성을 지닌 잡음이 있는 경우, 잡음구간임에도 불구하고 큰 값을 나타내는 단점이 존재한다. 이러한 단점을 해결하기 위해 새로운 소프트 문턱치 기법을 제안한다.

3.5 소프트 문턱치 음성구간 검출법

일련의 과정을 통해 계산된 웨이브렛 밴드 엔트로피 앙상블 분산 신호는 음성구간을 강조시켜 효과적으로 음성구간 검출에 활용할 수 있다. 하지만 잡음구간에서 음성과 비슷한 특성을 가진 잡음으로 인해

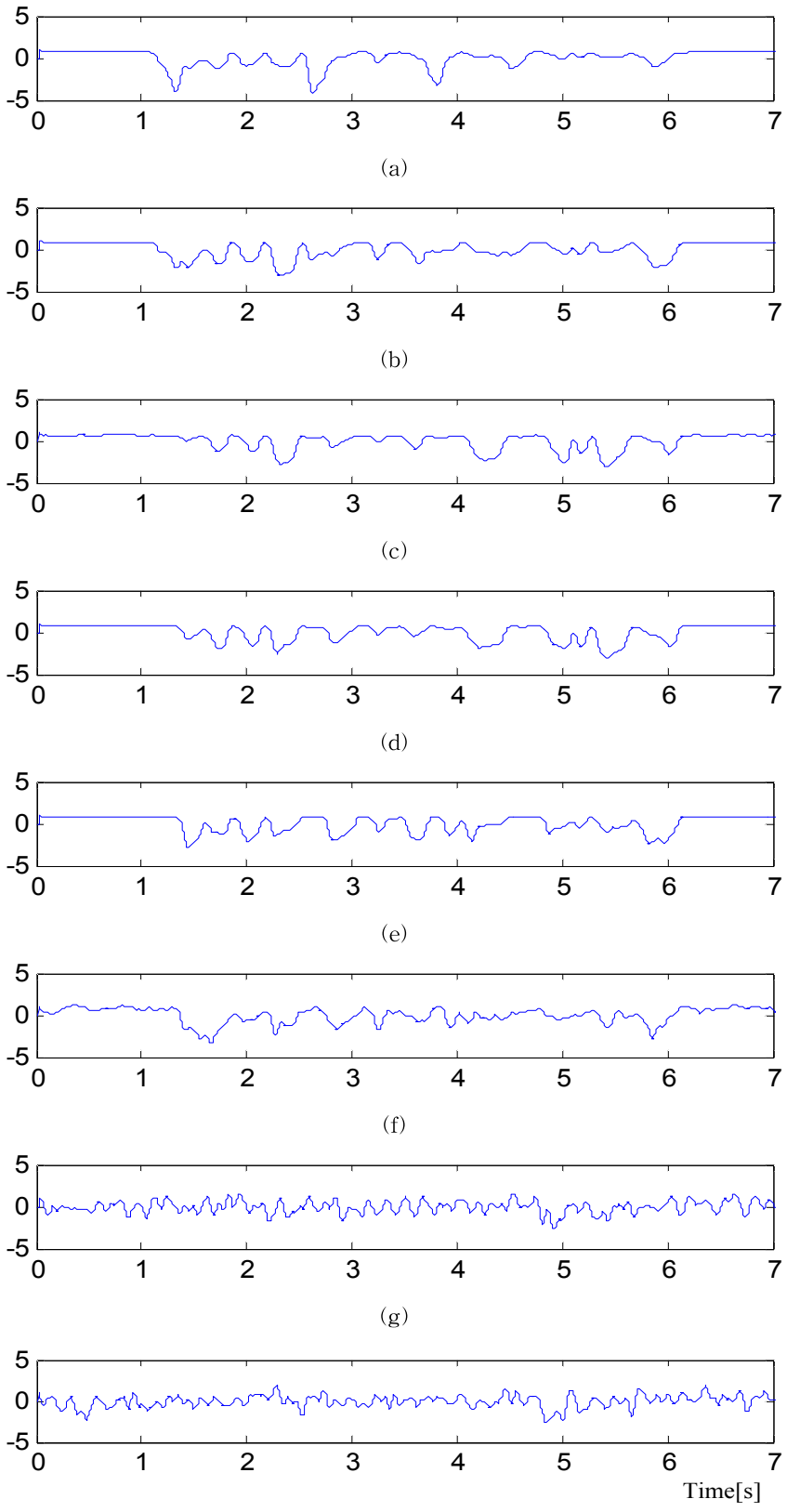

그림 5. 그림 4의 분해된 신호에 대한 웨이브렛 밴드 엔트로피 신호 (a) d1, (b) d2, (c) d3, (d) d4, (e) d5, (f) d6, (g) d7, (h) a7

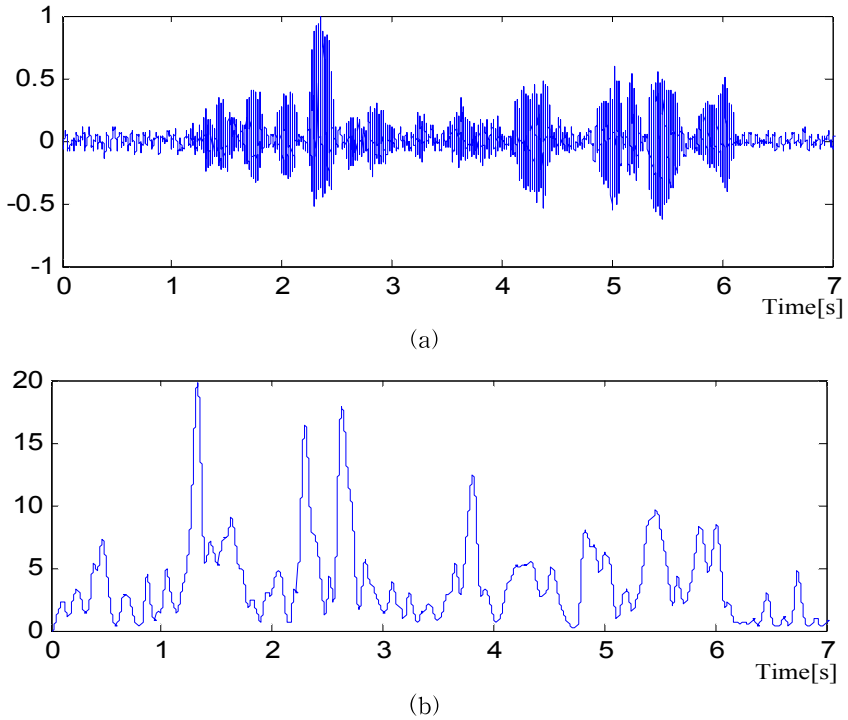

그림 6. 웨이브렛 밴드 엔트로피 앙상블 분석 (a) 자동차 잡음과 음성이 섞인 신호 (SNR 5 dB), (b) 웨이브렛 밴드 엔트로피 앙상블 분산

웨이브렛 밴드간의 분산이 큰 값이 나타날 수 있다. 이러한 부분을 해결하기위해서는 항상 같은 값을 문턱치로 정하여 사용하는 하드 문턱치 검출법을 사용하기보다는 주변 상황에 맞게 문턱치를 변화시키면서 음성구간을 검출하는 소프트 문턱치 검출법을 사용하여야 한다. 소프트 문턱치 검출법은 다음의 과정을 거친다. 식 (7)과 식 (8)을 이용하여 입력된 음성신호의 전체 엔트로피를 계산한다. 시간에 따라 계산된 신호를 $\rho_y(t)$라 하고 식 (12)에 따라 50 ms 윈도우(window)의 이동평균을 취한다.

$$\tilde{\rho}_y(t) = \frac{1}{N}\sum_{i=t-2/N}^{t+2/N} \rho_y(i) \tag{11}$$

$$T(t) = \alpha\tilde{\rho}_y(t) \tag{12}$$

$$VAD_s(t) = \begin{cases} \sigma_\epsilon(t) > T(t), & 1 \\ \sigma_\epsilon(t) \le T(t), & 0 \end{cases} \tag{13}$$

식 (11)을 통해 얻은 $\tilde{\rho}_y(t)$를 스케일 파라미터 α를 곱한 식 (12)로 소프트 문턱치를 결정한다. 입력된 신호의 처음 1초간은 음성이 없다고 가정하에서 실험을 수행하였다. 여기서 스케일 파라미터 α는 처음 1초간 잡음의 특성에 따라 문턱치와 음성구간 검출 신호 사이의 스케일을 맞춰주는 파라미터이다. 그리고 식 (13)을 통해 음성구간을 검출하였다.

4. 실험 결과 및 고찰

4.1 음성구간 검출 결과

제안한 알고리듬의 성능분석을 위한 실험 데이터로써 TIMIT 데이터베이스의 음성신호 샘플과 NOISEX-92의 잡음신호 샘플을 사용하였으며 이러한 데이터 샘플은 16kHz 샘플링레이트(sampling rate), 32비트(bit)의 해상도를 갖는다. 또한 다양한 SNR변화에 대한 알고리듬의 성능을 평가하기 위해 SNR을 0 dB, 5 dB, 10 dB, 15 dB로 구분하여 실험하였다. 제안한 알고리듬의 성능을 확인하기 위하여 일반적으로 음성구간 인식에 많이 쓰이는 MAE(maximum amplitude envelope) 알고리듬 및 Asgari[7]가 제안한 EVAD(entropy voice activity detection) 알고리듬과 비교하였다. 각각의 알고리듬에 대한 문턱치는 실험을 통해 가장 좋은 결과를 보여준 문턱치를 사용하였다. 이상의 과정을 통해 계산된 음성구간 검출 결과를 다양한 SNR 환경에서 실험하였으며 세 가지 음성구간 검출 알고리듬에 의해 검출된 음성구간을 그림 7에서 나타내었다.

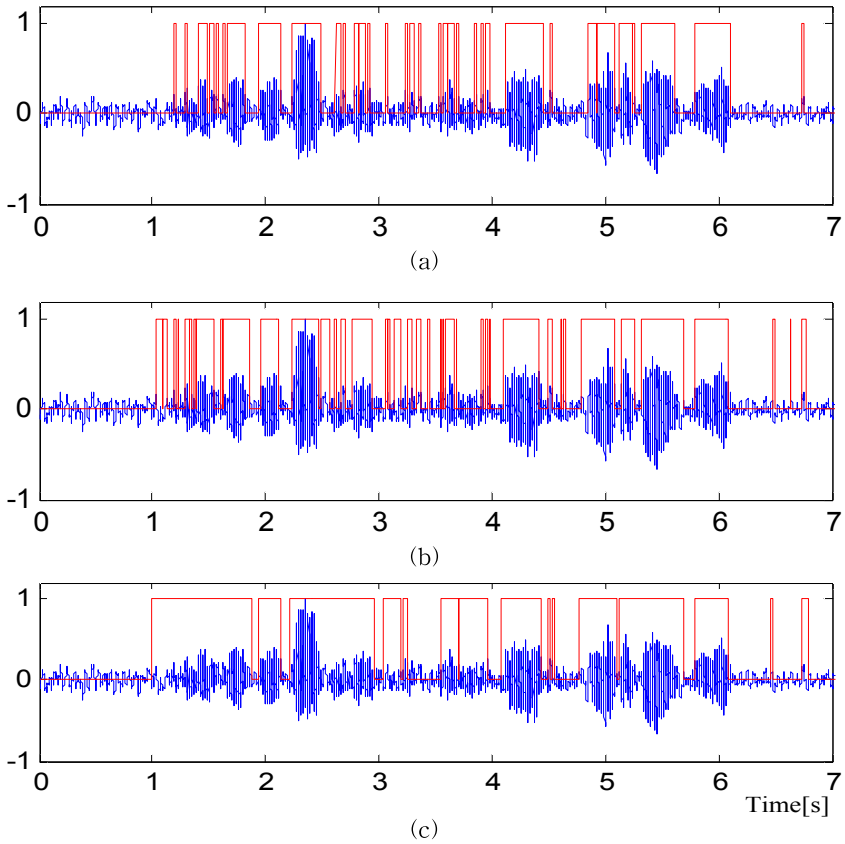

그림 7. SNR 0 dB에서 기존의 알고리듬들과 제안한 알고리듬을 이용해 검출한 음성구간 (a) MAE, (b) EVAD, (c) 제안한 알고리듬

그림 7은 잡음이 매우 강한 SNR 0 dB에서의 각 알고리듬에 대한 음성구간 검출 결과이다. MAE 알고리듬은 정해진 프레임의 최대값으로 포락선화(enveloping)하여 음성구간을 검출하는 방법으로서 잡음이 음성과 섞여 신호의 크기가 커지는 구간에서 잡음을 음성으로 잘못 인식하는 문제점이 있었다. 그림 7(a)는 MAE 알고리듬의 음성구간 검출 결과이다. 잡음의 크기가 큰 구간에서 잡음을 음성으로 인식하기도 하고 음성구간의 음성크기가 작아지는 구간에서 음성을 잡음으로 인식한 결과를 볼 수 있다. 전체적으로 음성구간이 작게 분할되어 음성구간이 정확하게 분리 되지 않은 결과를 나타내었다. 그림 7(b)는 EVAD 알고리듬의 음성구간 검출 결과로서 스펙트럼 영역에서의 엔트로피를 이용하여 그림 7(a)에서는 검출하지 못한 작은 크기의 음성구간도 일부 검출 하는 것을 볼 수 있다. 그러나 전체적으로 그림 7(a)의 결과와 마찬가지로 음성구간이 너무 작은 크기로 많이 분할되어 정확한 음성구간 검출을 하지 못하는 결과를 보였다. 그림 7(c)는 제안한 알고리듬의 음성구간 검출 결과이다. SNR이 낮은 상황임에도 불구하고 전체적으로 대부분의 음성구간을 정확하게 찾아내고 그림 7(a) 및 (b)와 달리 넓은 음성구간을 크게 분할하는 좋은 결과를 보였다. 하지만 잡음의 크기가 큰 상황으로 일부 잡음구간을 음성구간으로 잘못 인식하는 결과도 보였다.

그림 8은 SNR 5 dB에서의 음성구간 검출 결과로서 MAE 알고리듬 및 EVAD 알고리듬의 결과인 그림 8(a)와 (b)에서는 그림 7에 비해 검출 오류가 많이 줄어들었지만 여전히 많은 검출 오류가 나타나고 정확한 음성구간 검출을 하지 못하는 것을 볼 수 있다. 제안한 알고리듬의 결과인 그림 8(c)는 그림 7에서 나타나던 잡음구간에서의 검출 오류가 거의 사라지고 음성구간에서도 대부분의 구간을 정확히 검출한 것을 볼 수 있다.

그림 9는 10 dB에서의 음성구간 검출 결과이다. 그림 9(a) 및 (b)의 결과를 보면 이전의 SNR이 낮은 상황에서 보다 잡음영역에서 훨씬 좋은 결과를 보인다. 하지만 여전히 크기가 작은 음성구간을 제대로

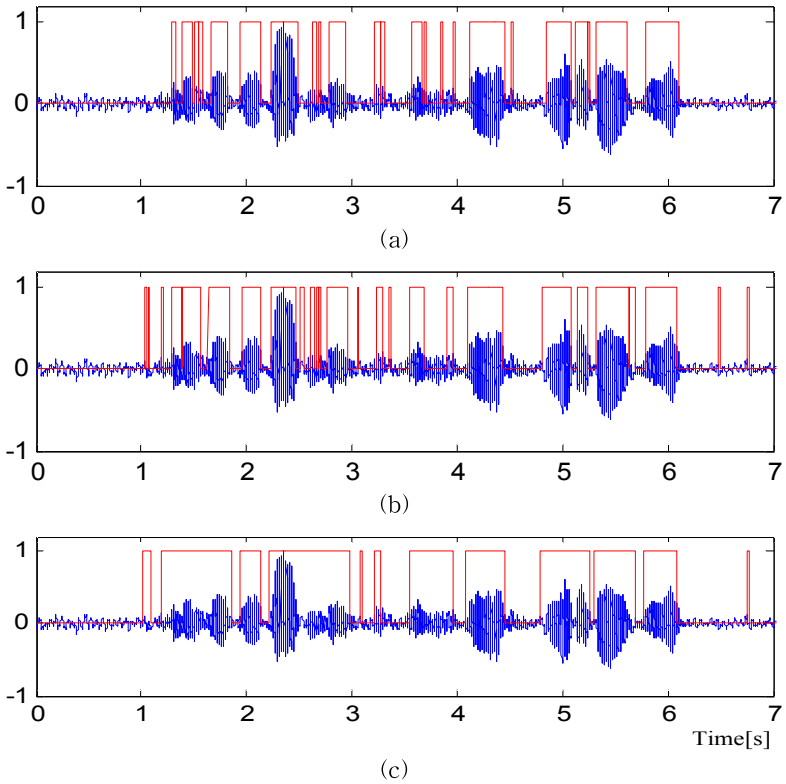

그림 8. SNR 5 dB에서 기존의 알고리듬과 제안한 알고리듬을 이용해 검출한 음성구간 (a) MAE, (b) EVAD, (c) 제안한 알고리듬

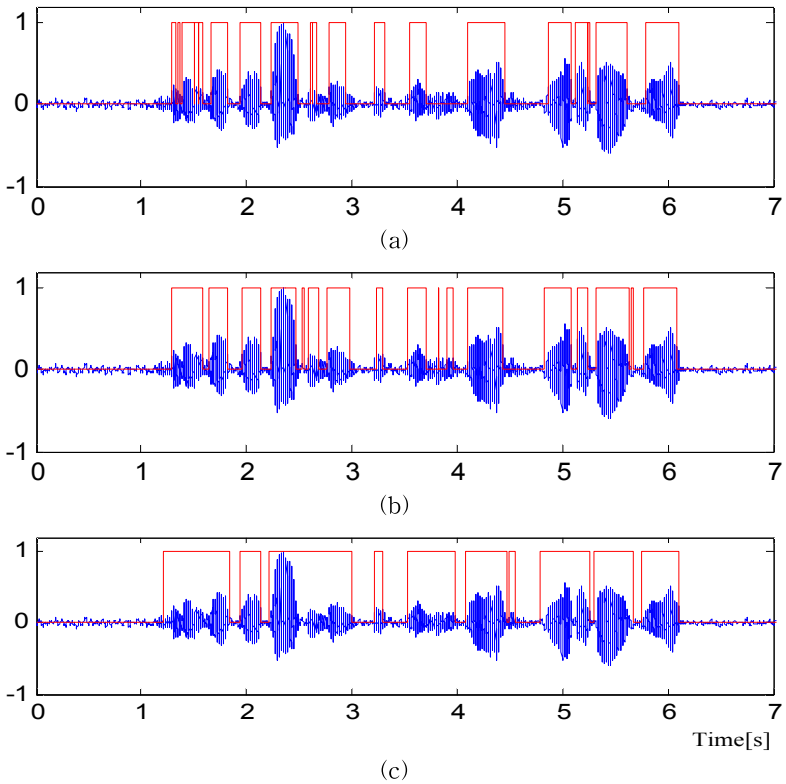

그림 9. SNR 10 dB에서 기존의 알고리듬과 제안한 알고리듬을 이용해 검출한 음성구간 (a) MAE, (b) EVAD, (c) 제안한 알고리듬

검출하지 못하는 문제가 나타나며 일부의 잡음구간을 음성구간으로 인식하는 오류가 나타난다. 반면, 그림 9(c)의 제안한 알고리듬의 결과는 거의 완벽하게 음성구간을 검출했으며 잡음구간을 음성구간으로 인식하는 오류도 거의 나타나지 않는 것을 볼 수 있다.

그림 10은 비교적 잡음이 적은 상황이라고 할 수 있는 SNR 15 dB의 환경이다. 세 알고리듬 모두 상당히 좋은 음성구간 검출 결과를 보이지만 그림 10(a)와 (b)에서는 일부 크기가 작은 음성신호가 있는 구간을 잡음으로 인식하는 오류를 보인다. 하지만 제안한 알고리듬의 결과인 그림 10(c)는 상당히 작은 크기의 음성구간도 찾아내며 완벽한 음성구간 검출 성능을 보였다. 이상의 실험 결과를 객관적인 성능으로 평가하기 위해 음성구간 적중률(pause hit rate, PHR)을 이용하였다. 여기서 PHR은 깨끗한 음성신호에서 수동으로 찾은 음성구간과 알고리듬을 통해 찾은 음성구간이 일치하는 정도를 나타낸 것이다. 식 (14)와 식 (15)에 따라 계산 되었다.

$$v(t)=\begin{cases}1, & \text{if} \quad VAD_m(t)=VAD_a(t)\\ 0, & otherwise\end{cases} \tag{14}$$

$$PHR=\frac{\sum_{t=1}^{L}(v(t))}{\sum_{t=1}^{L}(VAD_m(t))}\times 100 \tag{15}$$

여기서 VAD_m은 수동으로 찾은 음성구간을 나타내며 VAD_a은 알고리듬을 통해 찾은 음성구간을 나타낸다. 그리고 PHR은 백분율로 나타냈으며 높을수록 음성구간 검출 능력이 좋음을 나타낸다. 이러한 지표를 적용하여 제안한 알고리듬과 기존의 2가지 알고리듬의 음성구간 검출 성능을 표 1에서 보였다. MAE 알고리듬은 전체적으로 PHR이 50%대의 가장 좋지 않은 성능을 나타내었다. 그리고 EVAD 알고리듬과 제안한 알고리듬을 비교할 때 PHR에서 제안한 알고리듬이 기존의 EVAD 알고리듬보다 15~20% 이상 더 높은 적중률을 보여 매우 좋은 성능을 나타내었다. 특히 제안한 알고리듬은 SNR이 낮은 환경에서 높은 PHR을 유지하여 자동차 잡음이 심한 환경에서 사용할 수 있는 우수한 알고리듬으로 판단된다.

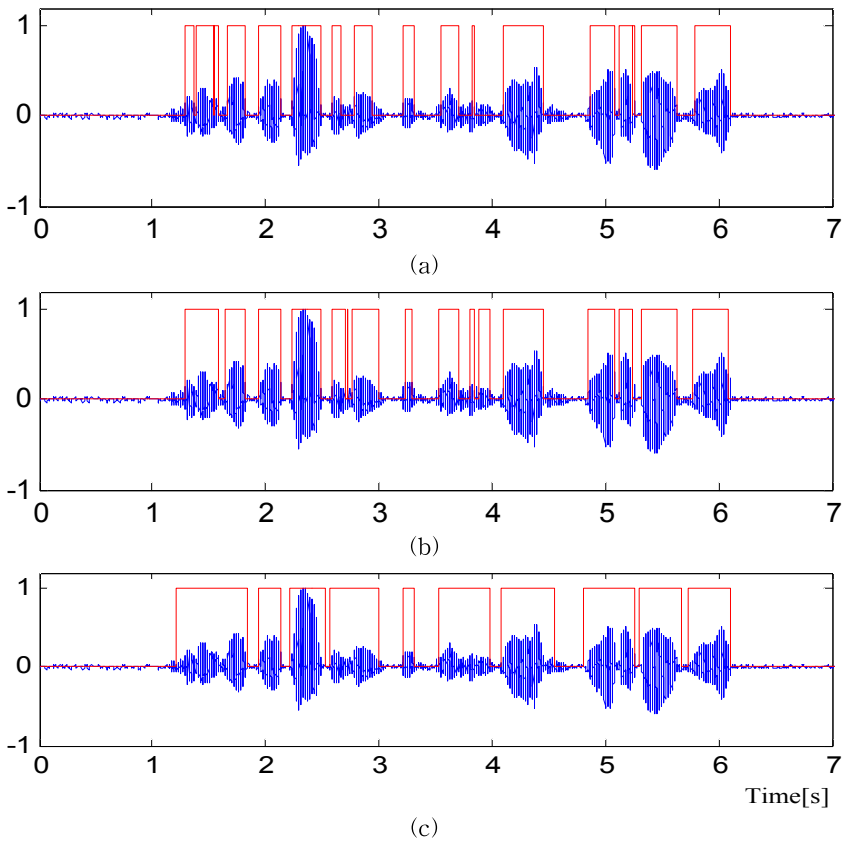

그림 10. SNR 15 dB에서 기존의 알고리듬과 제안한 알고리듬을 이용해 검출한 음성구간 (a) MAE, (b) EVAD, (c) 제안한 알고리듬

표 1. 제안한 알고리듬과 기존 음성구간 검출 알고리듬의 성능 비교 [%]

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

5. 결 론

본 논문에서는 새로운 웨이브렛 밴드 엔트로피 기반의 앙상블 분산과 소프트 문턱치를 이용한 음성검출 알고리듬을 제안하였다. 실험을 통해 제안한 음성검출 알고리듬이 기존의 알고리듬보다 정성적 및 정량적인 관점에서 모두 좋은 성능을 보였다. 제안한 웨이브렛 밴드 엔트로피 기반의 앙상블 분산은 웨이브렛 밴드별로 엔트로피의 차이가 큰 음성구간을 크게 증폭시키는 장점이 있지만 시간에 따른 변화가 심하고 음성과 비슷한 주파수 대역에 존재하는 자동차 잡음에서는 기존의 알고리듬과 마찬가지로 음성구간 검출이 어려운 단점을 가졌다. 이러한 단점을 극복하기 위해 제안한 소프트 문턱치 기법과 함께 사용함으로써 자동차 잡음환경에서의 음성구간 검출에서도 좋은 결과를 얻었다. 또한 높은 SNR환경에서 안정적이고 뛰어난 성능을 보일뿐 아니라 낮은 SNR환경에서는 더 좋은 성능을 발휘하는 것을 확인하였다.

제안한 음성구간 검출 알고리듬은 음성인식, 잡음제거 등의 분야에 적용할 수 있으며 특히, 음성통신의 전처리과정으로 활용된다면 음성 인식률 상승을 기대할 수 있다. 그리고 최근 각광받고 있는 스마트기기에서 음성명령이나 음성비서를 사용할 때 더욱 유용할 것으로 보인다. 특히 자동차 잡음이 심한 도로변에서 스마트기기를 사용하거나 스마트 자동차에 내장된 스마트장비에 음성명령을 내릴 때 자동차의 소음에 영향을 받지 않고 사용할 수 있는데 기여할 수 있을 것으로 기대된다.

참 고 문 헌

[1] L. Rabiner and B.H. Juang, *Fundmentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] D.G. Ha, S.J. Cho, G.G. Jin, and O.K. Shin, "Voice Activity Detection Based on Signal Energy and Entropy-difference in Noisy Environments," *Journal of the Korean Society of Marine Engineering*, Vol. 32, No. 5, pp. 768-774, 2008.

[3] J. Ramíirez, J.C. Segura, C. Beníitez, A. de la-Torre, and A. Rubio, "An Effective Subband OSF-based VAD with Noise Reduction for Robust Speech Recognition," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, No. 6, pp. 1119-1129, 2005.

[4] R. Gemello, F. Mana, and R. De Mori, "A Modified Ephraim-Malah Noise Suppression Rule for Automatic Speech Recognition," *Proc. ICASSP 2004*, Vol. 1, pp. 957-960, 2004.

[5] P. Teng and Y. Jia "Voice Activity Detection Via Noise Reducing using Non-Negative Sparse Coding," *IEEE Signal Processing Letters*, Vol. 20, Issue 5, pp. 475-478, 2013.

[6] Shi-Wen Deng and Ji-Qing Han, "Statistical Voice Activity Detection Based on Sparse Representation Over Learned Dictionary," *Digital Signal Processing*, Vol. 23, Issue 4, pp. 1228- 1232, 2013.

[7] M. Asgari, A. Sayadian, M. Farhadloo, and E.A. Mehrizi, "Voice Activity Detection using Entropy in Spectrum Domain," *Telecommunication Networks and Applications Conference*, pp. 407-410, 2008.

[8] C.E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 379-423, 1948.

[9] J. Ramirez, J.C. Segura, C.Benitez, L. Garcia, and A. Rubio, "Statistical Voice Activity Detection using a Multiple Observation Likelihood Ratio Test," *IEEE Signal Processing Letter*, Vol. 12, No. 10, pp. 689-692, 2005.

[10] H.K. Kim, S.W. Lee, and J.K. Hong, "Noise Reduction using Spectral Subtraction in the

Discrete Wavelet Transform Domain," *Journal of the Korea Multimedia Society*, Vol. 4, No. 4, pp. 306-315, 2001.

[11] J.I. Agbinya, "Discrete Wavelet Transform Techniques in Speech Processing," *IEEE TENCON. Digital Signal Processing Applications*, Vol. 2, pp. 514-519, 1996.

[12] S.H. Lee and D.H. Yoon, "EEG Signal Compression by Multi-scale Wavelets and Coherence Analysis and Denoising by Continuous Wavelets Transform," *Journal of the Institute of Electronics Engineers of Korea*, Vol. 41-SP, No. 3, pp. 221-229, 2004.

[13] S. Mallat and S. Zhong, "Caracterization of Signals from Multiscale Edges," *IEEE Trans. on Information Theory*, Vol. 38, No. 2, pp. 710-732, 1992.

[14] K.S. Bae, "Detecttion of Glottal Closure Instant for Voice Speech using Wavelet Transform," *Speech Sciences*, Vol. 7, No. 3, pp. 164-176, 2000.

[15] G.H. Lee, P.U. Kim, Y.J. Lee, and M.N. Kim, "Detection of the First and Second Heart Sound using Three-order Shannon Energy Difference," *Journal of the Korea Multimedia Society*, Vol. 14, No. 7, pp. 884-894, 2011.

이 기 현

2009년 8월 경북대학교 천문대기과학과(이학사)
2012년 2월 경북대학교 대학원 의용생체공학과(공학석사)
2012년 3월～현재 경북대학교 대학원 의용생체공학과 박사과정
관심분야 : 생체신호처리, 의용전자기기

이 윤 정

2003년 2월 경북대학교 전자전기공학부(공학사)
2005년 2월 경북대학교 대학원 의용생체공학과(공학석사)
2005년 3월～현재 경북대학교 대학원 의용생체공학과 박사과정
관심분야 : 생체신호처리, 의용전자기기

김 명 남

1988년 2월 경북대학교 전자공학과(공학사)
1990년 2월 경북대학교 대학원 전자공학과(공학석사)
1995년 2월 경북대학교 대학원 전자공학과(공학박사)
1996년～현재 경북대학교 의학전문대학원 의공학교실 주임교수
관심분야 : 생체신호처리시스템, 의학영상처리