한국음향학회지 제34권 제4호 pp. 310~315 (2015)
The Journal of the Acoustical Society of Korea Vol.34, No.4 (2015)
http://dx.doi.org/10.7776/ASK.2015.34.4.310

pISSN : 1225-4428
eISSN : 2287-3775

비정체성 잡음을 위한 SPD-TE 기반 계수형 음성 활동 탐지

A Parametric Voice Activity Detection Based on the SPD-TE for Nonstationary Noises

구본응†

(Boneung Koo†)

경기대학교 전자공학과

(Received Janyary 29, 2015; accepted April 7, 2015)

초 록: 본 논문에서는 비정체성(nonstationary) 잡음 환경을 위한 단일 채널 VAD(Voice Activity Detection) 알고리듬 제안하였다. VAD 판별을 위한 특징계수의 임계값은 과거 비음성 프레임들의 평균과 표준편차를 추산하여 적응적으로 갱신하였다. 특징계수로는 SPD-TE(Spectral Power Difference-Teager Energy)를 사용했는데, 이것은 WPD (Wavelet Packet Decomposition) 계수에 Teager 에너지를 적용한 것으로서 잡음에 강인한 것으로 보고된 바 있다. TIMIT 음성과 NOISEX-92 잡음을 사용하여 10 dB부터 -10 dB까지의 SNR에 대한 실험 결과, 제안된 알고리듬이 표준을 포함한 기존의 알고리듬과 비슷한 정확도를 보였다.

핵심용어: 음성 탐지, 비음성 탐지, 비정체성 잡음, 잡음 강인성, 단일 채널

ABSTRACT: A single channel VAD (Voice Activity Detection) algorithm for nonstationary noise environment is proposed in this paper. Threshold values of the feature parameter for VAD decision are updated adaptively based on estimates of means and standard deviations of past non-speech frames. The feature parameter, SPD-TE (Spectral Power Difference-Teager Energy), is obtained by applying the Teager energy to the WPD (Wavelet Packet Decomposition) coefficients. It was reported previously that the SPD-TE is robust to noise as a feature for VAD. Experimental results by using TIMIT speech and NOISEX-92 noise databases show that decision accuracy of the proposed algorithm is comparable to several typical VAD algorithms including standards for SNR values ranging from 10 to -10 dB.

Keywords: Voice activity detection, Speech pause detection, Nonstationary noise, Noise-robustness, Single channel
PACS numbers: 43.72.Ar

I. 서 론

음성 신호에 포함된 음성/비음성 구간을 판별하는 것을 음성활동탐지(VAD, Voice Activity Detection)라고 한다. VAD는 비음성 구간을 탐지하여 잡음의 스펙트럼 등 잡음 제거에 필요한 특성을 추출하는데 사용되고, 음성인식시스템의 인식률을 높이고, 보청기의 명료도를 높이고, 음성 코딩에서는 비음성 구간을 탐지하여 데이터 압축률을 높이는데 사용된다.

이러한 중요성 때문에 많은 연구가 이루어져 왔고[1] 대표적으로는 통계 모델에 기반한 알고리듬[2]과 국제 표준[3-5]이 있다. 그러나, 대부분의 VAD는 실제 잡음 환경에서는 정확도가 저하되는데, 그 이유는 잡음의 종류(백색 및 다양한 유색 잡음), 낮은 SNR, 잡음의 비정체성 등이다. 최근 십여 년 동안 비정체성 잡음에 대한 성능을 개선하기 위한 알고리듬이 제안되었다.[6-11] 이 알고리듬들의 핵심 요소는 잡음에 강인한 특징계수와 잡음 특성의 시간적 변화에 따라 특징계수의 임계값을 갱신하는 적응 알고리듬이다. 특징계수로는 프레임 에너지,[6,8] 스펙트럼,[7,10] 웨이블렛 계수[9] 등 에너지 기반 값들이 사용되었고,

†**Corresponding author:** Boneung Koo (bkoo@kgu.ac.kr)
Department of Electronic Engineering, Kyonggi University, 154-42 Gwanggyosan-Ro, Yeongtong-Gu, Suwon 443-760, Republic of Korea
(Tel: 82-31-249-9798, Fax: 82-31-244-6300)

두 가지 이상의 특징계수와 패턴인식 기법이 사용되기도 한다.[11] 에너지 기반 특징계수들의 공통적인 단점은 SNR이 적어질수록 변별력이 저하되는 것이다. 또한, 임계값 갱신 알고리듬이 경험적이고 추가되는 상수의 결정 과정이 실험적이어서 다양한 잡음 환경에 적용하기 어렵다.

Ramirez는 긴 시간에 걸친 음성과 잡음의 스펙트럼의 차이를 특징계수로 사용하여 잡음의 비정체성을 정면으로 다루었다.[6] 그러나, 현실적으로 비정체성 잡음의 평균 스펙트럼을 갱신하기 어렵고, 긴 분석 창을 사용하므로 필연적으로 시간 지연을 초래한다. Ghosh는 LTSV(Long-Term Signal Variability)라는 특징계수를 사용하여 몇 가지 국제표준보다 나은 정확도를 보였다.[10] 그러나, 알고리듬이 복잡하고 판별에 사용된 voting 방식만으로도 0.3 s의 지연 시간이 생긴다.

본 연구에서는 매 프레임마다 임계값을 갱신하는 적응 알고리듬을 제안하였다. 특징계수의 평균과 표준편차를 사용하므로 계수형 VAD라고 칭한다. II장에 제안된 알고리듬을 설명하고, III장에 실험 결과를, IV장에 결론을 제시하였다.

II. 계수형 VAD 알고리듬

특징계수로는 SPD(Spectral Power Distance)[9]에 TE (Teager Energy)를 적용한 SPD-TE[12]를 사용했는데, 이것은 비정체성 잡음에 대하여 기존의 통계 모델 기반 정체성 VAD[2]보다 높은 정확도를 보인 바 있다.[12] SPD-TE를 구하는 과정은 다음과 같다.

$$D(i)=\left|\frac{1}{N_a}\sum_{n=1}^{N_a}E_{m,i}(n)-\frac{1}{N_b}\sum_{n=N_a+1}^{N}E_{m,i}(n)\right|, \quad (1)$$

여기서 $E_{m,i}(n)$는 입력 신호의 2-채널 WPD(Wavelet Packet Decomposition) 계수 $X_{m,i}(n)$의 TE

$$E_{m,i}(n)=X_{m,i}^2(n)-X_{m,i}(n+1)X_{m,i}(n-1), \quad (2)$$

이고 N은 frame length, m은 scale index, i는 frame index, $N_a=N/2^m$와 $N_b=N-N_a$는 저주파수 대역 및 고주파수 대역 WPD 계수의 개수이다. $f_s=8$ kHz일 때 저대역과 고대역의 경계는 2 kHz이다.

$$D_w(i)=D(i)\left[\frac{1}{2}+\frac{16}{\log(2)}\log\left(1+2\sum_{k=1}^{N}s_i^2(n)\right)\right], \quad (3)$$

$$D_c(i)=\frac{1-e^{-2D_w(i)}}{1+e^{-2D_w(i)}}. \quad (4)$$

특징계수 SPD-TE는

$$\gamma(i)=D_c(i)*h(i), \quad (5)$$

$$H(z)=\frac{1}{1-a_1z^{-1}}, \quad a_1=0.65, \quad (6)$$

이다. 자세한 내용은 Reference [12]에 있다.

특징계수의 임계값은 다음과 같이 설정하였다.

$$\theta(i)=\mu_N(i)+p(i)\,\sigma_N(i), \quad (7)$$

여기서 $\mu_N(i)$, $\sigma_N(i)$는 각각 이전 1 s에 해당하는 잡음 프레임들의 $\gamma(i)$의 평균과 표준편차이다. 임계값 결정계수 $p(i)$로 상수 p를 사용해도 $\mu_N(i)$, $\sigma_N(i)$의 변화에 따라 임계값이 갱신된다. 이전 연구에서는 p 값을 가변하여 구한 ROC(Receiver Operating Characteristics) 곡선으로 성능을 평가하였다.[12] 본 연구에서는 매 프레임마다 $p(i)$ 값을 아래와 같은 적응 알고리듬을 사용하여 구한다.

예비 실험 결과, 음성 구간과 비음성 구간의 특징계수들의 의미있는 확률 모델은 구할 수 없었다. 그 대신, 음성 구간과 비음성 구간의 평균과 표준편차를 구하여 $\mu_S+k\,\sigma_S$와 $\mu_N+k\,\sigma_N$을 그려보니 잡음에 따라 차이는 있지만 대체로 SNR이 낮을 때는 상당한 중첩이 있었다. 여기서, μ_S, σ_S은 음성, μ_N, σ_N은 잡음 프레임들의 특징계수의 평균과 표준편차이고, $\mu_S>\mu_N$이다. 가우시안이면 $k=3$일 때 99.7 %를 포함한다. 임계값은 μ_S과 μ_N 사이에 설정하는데 그 간격이 클수록 임계값 결정계수 $p(i)$ 값을 크게 하여

더 많은 잡음을 포함하게 한다. 그 간격의 크기를 가늠하는 계수로 $r(i)=\{\mu_S(i)-\mu_N(i)\}/\sigma_N(i)$ 를 사용하고, 임계값 결정계수는 다음 식으로 구한다.

$$p(i)=\begin{cases} p_{\min} & , \quad r(i)<r_1 \\ p_{\min}+\left(\dfrac{r(i)_{\mathrm{dB}}-r_{1_{\mathrm{dB}}}}{r_{2_{\mathrm{dB}}}-r_{1_{\mathrm{dB}}}}\right)(p_{\max}-p_{\min}) & , \quad r_1<r(i)<r_2 \\ p_{\max} & , \quad \text{otherwise} \end{cases} \tag{8}$$

상수 $\{r_1, r_2, p_{\min}, p_{\max}\}$는 실험적으로 결정한다. 음성과 잡음의 특징계수의 평균과 표준편차를 갱신하기 위하여 음성과 잡음의 특징계수 값을 저장하는 두 개의 버퍼를 사용한다. 각 버퍼의 길이는 1 s,[10] frame advance 10 ms이므로 100개 프레임의 특징계수 값이 저장된다. 처음 1 s 동안은 비음성 신호라고 가정, 잡음 버퍼를 채우고 $p(i)$ 의 초기치로는 $p_o=3$ (실험치)을 사용하고, 그 이후에는 음성 및 잡음 버퍼를 갱신하며 임계값을 구한다.

알고리듬을 요약하면 다음과 같다.

1) Pre-emphasis: 70 Hz HPF(5-th Butterworth).

2) Window: 25 ms Hanning, 10ms advance.

3) WPD: 2 subbands, DB4(4-th order Daubechies).

4) 음성, 잡음 버퍼로부터 $\theta(i)$ 계산: Eq. (7), (8).

5) 특징계수 SPD-TE $\gamma(i)$ 계산: Eq. (1)-(6).

6) $\mathrm{vad}(i)=\begin{cases}1('\mathrm{speech}'), & \mathrm{if}\ \gamma(i)>\theta(i) \\ 0('\mathrm{nonspeech}'), & \mathrm{otherwise}\end{cases}$

7) $\mathrm{vad}(i)$ 값에 따라 음성 또는 잡음 버퍼 갱신.

8) Hangover: buffer length N=5[5].

9) 프레임 단위로 반복.

III. 실험 결과

TIMIT speech corpus[13]의 Core Test set는 8개 언어 지역(Dialect Region)별로 3인(남성 2, 여성 1)의 화자가 녹음한 단문(text) 8개씩, 도합 192개 단문으로 구성되어 있다. 단문의 길이는 2내지 4 s 가량인데, 화자 1인당 단문 8개를 앞, 뒤, 중간에 2 s의 침묵(silence) 구간을 삽입하여 하나의 장문(long sentence)으로 편집하여 실험에 사용하였다. 이러한 장문은 이전 연구에서 사용된 바 있다.[10,12] 장문 1개의 길이는 대략 40 s, 음성 구간은 대략 50 %이고, 본 실험에서 사용된 Test set는 장문 24개이다. 또한, 알고리듬 최적화와 성능 검사를 위한 음성 데이터를 분리하기 위하여 TIMIT Training set으로부터 Core Test set와 같은 포맷으로 192개의 단문을 추출, 24개의 장문으로 Training set을 편집하여 예비 실험(II절) 및 상수 최적화에 사용하였다.

VAD 비교 기준인 음성 구간의 index는 TIMIT의 *.phn 파일에 주어진 시작, 끝, pause 구간의 index를 수정하여 사용했고, 잡음은 NOISEX-92[14]의 잡음 중 다섯 가지(Volvo, Babble, White, Pink, Factory2)를 사용하였다. 음성과 잡음은 $f_s=8$ kHz로 미리 down-sampling하였다.

장문의 길이 40여 s 동안에 NOISEX-92 잡음은 느리게 변하는 비정체성이다. 입력 신호는 전체적인 SNR이 10, 5, 0, -5, -10 dB가 되도록 잡음 신호 크기를 조절하였다. 따라서, 한 가지 잡음에 대한 실험 세트는 24*5=120개, 모든 잡음과 모든 SNR에 대한 전체 실험 세트는 120*5=600개 장문으로 구성된다.

Training set을 사용한 알고리듬 및 상수 최적화를 통하여 $p_o=3$, $[r_1, r_2, p_{\min}, p_{\max}]=[4,\ 100,\ 1,\ 7]$ 로 설정하였다.

Ghosh[10]와 비교하기 위하여 다음과 같은 성능 평가 기준을 사용하였다.

1) CO(Correct): '1'→'1', '0'→'0 정확도.

2) TR(True Rejection): '1'→'0' 오류율.

3) FA(False Acceptance): '0'→'1' 오류율.

'1'=speech, '0'=non-speech, TR=FEC+MSC, FA=OVER+NDS, CO=1-(TR+FA)이고, FEC(Front End Clipping), MSC(Mid Speech Clipping), OVER(Carry over), NDS (Noise Detected as Speech)[10]는 오류 속성을 세분화한 것으로서, 0에 가까울수록 좋다. 따라서, TR과 FA는 0에 가까울수록, CO는 1에 가까울수록 좋다.

VAD 동작의 예를 Fig. 1에 보였다. 첫 번째 그림에는 FDHC0 음성 파형과 음성 구간, 두 번째 그림은 -10 dB Volvo 잡음이 더해진 신호, 세 번째 그림은

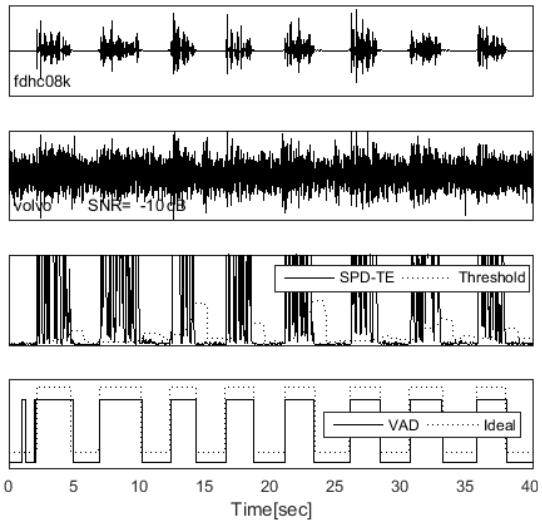

Fig. 1. Example plots: clean speech (top), −10 dB Volvo noisy speech (second), SPD-TE and threshold (third), and VAD results (bottom).

SPD-TE(실선)와 임계값(점선), 맨 아래 그림은 VAD 판별 결과(실선)로서 이상적인 기준(점선)과 거의 같다. 점수는 (CO, TR, FA) = (.9717, .0094, .0189)이다.

Table 1은 SPD-TE와 적응 알고리듬을 사용하여 본 연구에서 제안한 VAD의 Test set에 대한 실험 결과이다. CO 점수에서 Volvo(자동차) 잡음에 대해서는 모든 SNR에서 0.94의 높은 정확도를 보였고, 음성과 유사한 Babble 잡음에 대해서는 SNR 감소에 따라 점수가 급격히 낮아지고 있으나 다른 잡음들의 경우 0 dB까지 90 % 이상의 정확도를 유지하고 있다. SNR 평균은 특정 잡음에서 모든 SNR에 대한 평균 점수로서 Volvo에 대해서 가장 우수하고 Babble에 대해서 가장 낮았다. 잡음 평균은 특정 SNR에서 모든 잡음에 대한 평균값으로서, SNR 저하에 따라 점수가 약간씩 낮아지다가 -10 dB일 때 급격히 낮아졌다. TR 점수는 0에 가까울수록 좋은데, 역시 Volvo가 가장 좋고, Babble이 가장 나쁘다. FA 점수도 비슷한 경향을 보이고 있다. 모든 SNR과 모든 잡음에 대한 전체 Test set의 평균 점수는 (CO, TR, FA) = (.8766, .0712, .0523)이다. TR과 FA가 어느 한 쪽으로 치우치지 않고 비슷한 값을 보이고 있는 것은 바람직한 현상이다. 알고리듬 최적화는 이 전체 평균 점수를 기준으로 Training set을 사용하여 이루어졌다.

Ghosh와 Narayanan는 Reference [10]에서 LTSV를 제안하고, G.729, AMR1(=AMR-VAD1),[4] AMR2(=AMR-VAD2)[4] 국제표준과 성능을 비교했는데, 그 중 월등한 성능을 보인 것은 LTSV와 AMR2이다. Fig. 2에 LTSV, AMR2 및 SPD-TE의 SNR 평균 점수를 잡음 별로 보였다. 세 개의 막대 그래프 중에서 맨 위가 CO, 가운데가 TR, 맨 아래가 FA이다. 붙어 있는 세 개의 막대 중에서 왼쪽이 AMR2, 가운데가 LTSV, 오른쪽이 SPD-TE이다. CO에서 Volvo의 경우 AMR2가 가장 좋고 SPD-TE가 LTSV보다 약간 우세하다. Babble의 경우에는 AMR2, LTSV, SPD-TE의 순서이고, White의 경우에는 LTSV가 가장 우세, SPD-TE가 AMR2보다 약간 우세하였다. Pink의 경우 AMR2, SPD-TE, LTSV의 순서이고, Factory2의 경우, AMR2, LTSV, SPD-TE의 순서이다. SPD-TE를 기준으로 LTSV와 비교해보면 Volvo와 Pink에 대해서 우세했고, Factory2에 대해서는 비슷했다. SPD-TE를 기준으로 AMR2와 비교해보면 White에 대해서 약간 우세, Volvo, Pink, Factory2에 대해서는 비슷했다. TR 점수는 White와 Pink에 대해서 SPD-TE가 AMR2보다 좋았으나, LTSV보다는 모든 잡음에 대해서 미치지 못 했다. FA점수는 Volvo, Pink, Factory2에 대해서 SPD-TE가 LTSV보다 좋았으나, 모든 잡음에 대해서 AMR2에는 미치지 못했다. 그러나, 여기서 한 가지 주목할 것은 AMR2와 LTSV에 비하여 SPD-TE가 TR과 FA의 편차가 적다는 점이다. 이것은 '1'→'0' 오류율과 '0'→'1' 오류율이 비슷한 것을 의미하며 일반적인 VAD의 바람직한 속성이라고 할 수 있다.

Fig. 3에는 −10 dB SNR에 대한 평균 점수를 잡음 별로 보였다. CO 점수는 Volvo, Pink, Factory2에 대하여 다른 VAD들과 대등하고, White에 대해서는 AMR2보다 우세하고, Babble에 대해서는 가장 낮았다. TR 점수는 White, Pink에 대하여 AMR2보다 우세, Factory2에 대해서는 다른 것들과 대등, Babble에 대하여 가장 열등하였다. FA 점수는 모든 잡음에[대하여 AMR2가 가장 우수하였고, SPD-TE가 LTSV보다 Pink에 대하여 우세, Babble에 대하여비슷, AMR2와는 Volvo에 대하여 비슷하였다.

비교 결과를 요약하면 다음과 같다.

1) SPD-TE는 광범위한 SNR(-10에서 10 dB)에 대하

Table 1. VAD scores for the SPD-TE adaptive VAD.

Score	Noise	10 dB	5 dB	0 dB	-5 dB	-10 dB	SNR Average
CO	Volvo	0.9347	0.9446	0.9526	0.9530	0.9388	0.9447
	Babble	0.9257	0.8820	0.8251	0.6954	0.5281	0.7713
	White	0.9619	0.9487	0.9059	0.8648	0.7484	0.8859
	Pink	0.9608	0.9520	0.9096	0.8745	0.7331	0.8860
	Factory2	0.9351	0.9466	0.9287	0.8673	0.7968	0.8949
	Noise Average	0.9436	0.9348	0.9044	0.8510	0.7490	0.8766
TR	Volvo	0.0005	0.0015	0.0032	0.0103	0.0365	0.0104
	Babble	0.0394	0.0857	0.1081	0.1061	0.3188	0.1316
	White	0.0096	0.0367	0.0863	0.1287	0.1560	0.0835
	Pink	0.0082	0.0306	0.0807	0.1174	0.1452	0.0764
	Factory2	0.0053	0.0174	0.0507	0.0904	0.1060	0.0540
	Noise Average	0.0126	0.0344	0.0658	0.0906	0.1525	0.0712
FA	Volvo	0.0648	0.0539	0.0442	0.0367	0.0247	0.0449
	Babble	0.0349	0.0323	0.0668	0.1985	0.1531	0.0971
	White	0.0285	0.0146	0.0078	0.0065	0.0956	0.0306
	Pink	0.0310	0.0174	0.0097	0.0081	0.1217	0.0376
	Factory2	0.0596	0.0360	0.0206	0.0423	0.0972	0.0511
	Noise Average	0.0438	0.0308	0.0298	0.0584	0.0985	0.0523

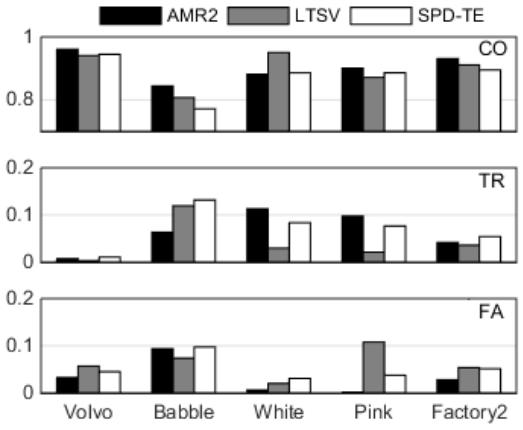

Fig. 2. Scores for average of all SNRs. From top to bottom: CO, TR, FA.

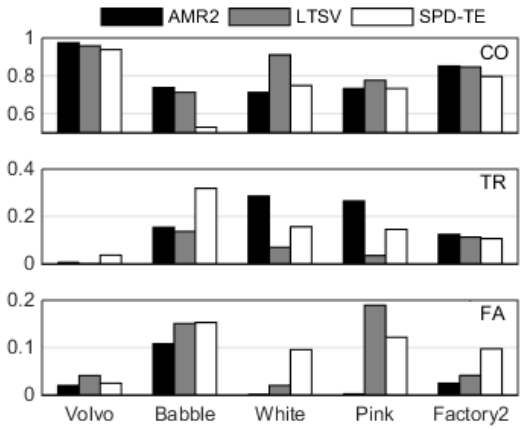

Fig. 3. Scores for SNR=-10 dB. From top to bottom: CO, TR, FA.

여 잡음 종류에 따라 AMR2, LTSV에 버금가거나 대등한 수준의 정확도를 보였다.

2) 음성으로 이루어진 Babble 잡음에 대하여 SPD-TE의 점수가 가장 낮았는데, 그 이유는 AMR2와 LTSV에 비하여 상대적으로 음성의 특징을 가장 많이 내포하고 있기 때문인 것으로 보인다.

3) 오류율에 있어서 TR과 FA점수가 AMR2, LTSV보다 균형 잡혀 있다.

IV. 결 론

본 논문에서는 음성과 잡음의 평균과 표준편차에 기반한 임계값 갱신 알고리듬을 제안하였다. 특징계수로는 잡음에 강인한 것으로 보고된 SPD-TE를 사용하여 비정체성 잡음에 강인한 VAD 알고리듬을 제안하였다. TIMIT 음성과 NOISEX-92 잡음을 사용한 실험 결과, SNR 0 dB에서도 평균 90 % 이상의 정확도

를 보였고, 기존의 대표적인 비정체성 VAD에 버금가는 정확도를 보였다. 정확도에서 비교 대상을 능가하지는 못했지만, 복잡도나 시간 지연 면에서 유리하다. 또한 -5 dB, -10 dB와 같이 낮은 SNR에서의 실험 결과를 제시한 논문이 드물다는 것을 감안하면 의미있는 결과라고 할 수 있다.

잡음의 비정체성의 속성에 따라 적응 알고리듬의 성능은 가변적일 수 밖에 없다. 실제 환경에서 사용하려면 다양한 종류의 잡음과 SNR의 변화 속에서도 높은 정확도를 유지하도록 VAD 알고리듬이 개선되어야 하고, 객관적인 비교 검증을 위한 비정체성 잡음의 표준 데이터베이스도 필요하다.

감사의 글

본 연구는 2014학년도 경기대학교 학술연구비(일반연구과제) 지원에 의하여 수행되었음.

References

1. P. C. Loizou, *Speech Enhancement* (CRC Press, Boca Raton, 2007), pp. 309-400.
2. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE Signal Process. Lett. **16**, 1–3 (1999).
3. ITU, *A silence compression scheme for G.729 optimized for terminals conforming to recommendation V.70, ITU-T Recommendation G.729-Annex B* (1996).
4. ETSI EN 301 708 V7.1.1(1999-12), *Digital cellular telecommunications system(Phase 2+); VAD for AMR speech traffic channels; General Description (GSM 06.94 version 7.1.1 Release 1998)*, 13-14 (1999).
5. ETSI ES 202 050, Ver. 1.1.5(2007-01), *Speech Processing, Transmission and Quality Aspects(STQ); Distributed Speech Recognition; Advanced front-end feature extraction algorithm; Compression algorithms, Annex A.3 Stage 2-VAD Logic*, 42-43 (2007).
6. J. Ramirez, J. C. Segura, C. Benitez, A. Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," Speech Commun. **42**, 271-287 (2004).
7. A. Davis, S. Nordholm, and R. Togneri, "Statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold," IEEE Trans. Audio, Speech and Lang. Processing **14**, 412-414 (2006).
8. G. Evangelopoulos and P. Maragos, "Multiband modulation energy tracking for noisy speech detection," IEEE Trans. Audio, Speech and Lang. Processing **14**, 2024-2038 (2006).
9. T. V. Pham and T. T. Chien, "Reliable voice activity detection algorithm under adverse environments," in Proc. IEEE Int. Conf. Commun. Electronics, 218-223 (2008).
10. P. K. Ghosh and S. Narayanan, "Robust voice activity detection using long-term signal variability," IEEE Trans. Audio, Speech and Lang. Processing **19**, 600-613 (2011).
11. E. Chuangsuwanich and J. Glass, "Robust voice activity detector for real world application using harmonicity and modulation frequency," in Proc. Interspeech, 2645-2648 (2011).
12. B. Koo, "A single channel voice activity detection for noisy environments using wavelet packet decomposition and Teager energy" (in Korean), J. Acoust. Soc. Kr. **33**, 139-145 (2014).
13. J. Garofolo, "TIMIT acoustic-phonetic continuous speech corpus," LDC93S1, Linguistic Data Consortium, Philadelphia, 1993.
14. A. Varga and H. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: An additive noise on speech recognition systems," Speech Commun. **12**, 247-251 (1993).

저자 약력

▸ 구 본 응 (Boneung Koo)

1975년 2월: 서울대학교 공업교육학과 전자공학전공 학사
1977년 1월~1982년 6월: 한국원자력연구소 연구원
1984년 12월: Texas A&M University, Dept. of Electrical Engineering, M.S.
1988년 12월: Texas A&M University, Dept. of Electrical Engineering, Ph.D.
1989년 3월 ~ 현재: 경기대학교 교수