한국음향학회지 제33권 제2호 pp. 139~145 (2014)
The Journal of the Acoustical Society of Korea Vol.33, No.2 (2014)
http://dx.doi.org/10.7776/ASK.2014.33.2.139

pISSN : 1225-4428
eISSN : 2287-3775

웨이블렛 패킷 변환과 Teager 에너지를 이용한 잡음 환경에서의 단일 채널 음성 판별

A Single Channel Voice Activity Detection for Noisy Environments Using Wavelet Packet Decomposition and Teager Energy

구본응†

(Boneung Koo†)

경기대학교 전자공학과

(접수일자: 2013년 12월 6일; 채택일자: 2014년 1월 24일)

초 록: 본 논문에서는 WPD (Wavelet Packet Decomposition) 계수에 Teager 에너지를 적용한 특징 계수를 임계값 알고리듬에 적용하여 잡음에 강인한 VAD 알고리듬을 제안하였다. 임계값은 비음성 구간의 평균과 표준편차를 추산하여 설정하였다. TIMIT 음성과 NOISEX 잡음 데이터베이스를 사용한 실험 결과, 제안된 알고리듬이 기존의 대표적인 비교 대상 알고리듬보다 우수함을 보였다. 정확도는 SNR 10 dB부터 -10 dB까지 ROC (Receiver Operating Characteristics) 곡선을 사용하여 비교하였다.

핵심용어: 음성 탐지, 비음성 탐지, Teager 에너지, 웨이블렛 패킷 변환, 잡음 강인성, 단일 채널

ABSTRACT: In this paper, a feature parameter is obtained by applying the Teager energy to the WPD(Wavelet Packet Decomposition) coefficients. The threshold value is obtained based on means and standard deviations of nonspeech frames. Experimental results by using TIMIT speech and NOISEX-92 noise databases show that the proposed algorithm is superior to the typical VAD algorithm. The ROC(Receiver Operating Characteristics) curves are used to compare performance of VAD's for SNR values of ranging from 10 to -10 dB.

Keywords: Voice activity detection, Speech pause detection, Teager energy, Wavelet packet decomposition, Noise-robustness, Single channel

PACS numbers: 43.72.Ar

I. 서 론

VAD(Voice Activity Detection)는 음성신호의 음성/비음성 구간 판별을 위한 것으로서, 음성인식, 잡음제거, 부호화 등에 사용된다. VAD는 입력 신호의 수에 따라 단일 채널 알고리듬과 다중 채널 알고리듬으로 구분할 수 있는데, 마이크를 여러 개 사용하는 다중 채널의 경우에는 성능 향상과 함께 시스템 복잡도가 증가한다. 단일 채널 방식은 일반적으로 SNR(신호 대 잡음비)이 클 경우에는 에너지와 영교차율 등 간단한 특징계수와 임계값 알고리듬으로 정확한 판별이 가능하지만 SNR이 작아질수록 정확도가 저하되고, 또한, 잡음 특성이 시간에 따라 변하는 비정체성(nonstationary)인 경우에는 정확한 판별이 매우 어렵다.[1,2]

본 논문은 정체성(stationary) 잡음에 강인한 단일 채널 VAD 알고리듬에 관한 것이다. 이러한 알고리듬의 핵심 요소는 특징계수와 판별 알고리듬이다. 특징계수는 음성과 배경 잡음을 구분해야 하는데, 음성은 물론이고 잡음도 실제 환경에서는 그 특성과 종류가 매우 다양하여 간단한 임계값 알고리듬부터 복잡한 패턴인식 기반 알고리듬까지 여러 가지 방법이 제안되었다.[3] 본 논문에서는 임계값 알고리듬을

†**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)

사용하였다.

잡음에 강인한 특징계수로는 에너지, 상관계수, 켑스트럼, 스펙트럼 엔트로피, 웨이블렛 변환 계수 및 이들을 응용한 다양한 계수들이 시도되었다.[4] 이러한 특징계수 기반 알고리듬들은 대부분 경험과 실험에 기반한 알고리듬으로서 SNR과 스펙트럼 변화 등 주변 상황에 따라 성능이 가변적이다. 성능 향상을 위하여 두 가지 이상의 계수를 사용하면 그에 따라 판별 알고리듬은 더욱 복잡해진다.[5,6] 이에 반하여 통계 모델에 기반한 알고리듬은[7] 실제 신호가 가정된 모델과 맞지 않을 개연성에도 불구하고, 빠르고 우수한 성능을 보임으로써 모든 VAD의 비교 대상으로 참조되고 있다. 몇 가지 국제 표준과 최근의 연구 동향은 참고문헌[8-14]에 있다.

TE(Teager Energy)는 에너지보다 잡음에 강인한 것으로 알려져 있다. 본 논문에서는 웨이블렛 계수에 TE를 적용하여 특징계수를 구하였다. II절에서는 특징계수와 임계값 구하는 과정을 설명하고, III절에는 실험 결과를, IV절에는 결론을 제시하였다.

II. 특징계수와 임계값 알고리듬

Time index를 n, 음성 신호를 $x(n)$, 잡음을 $d(n)$라 하면, 잡음이 더해진 신호는 $s(n)=x(n)+d(n)$이고 이 신호의 TE는 다음과 같이 정의된다.[15]

$$\Psi_s(n)=s^2(n)-s(n+1)s(n-1). \quad (1)$$

TE가 잡음에 강인한 것으로 알려지면서 음성인식, 잡음감쇠, 음성판별 등에 도입되었다.[13, 16, 17] 그러나, 자동차 잡음에 대한 실험 결과, 단일 대역에서는 잡음에 대한 강인함이 통상적인 STE(Short-Time Energy)보다 약간 낮거나 비슷한 수준이지만, subband에 기반한 경우에는 TE가 STE보다 훨씬 더 잡음에 강인하였다.[18]

Pham과 Chien[13]은 저주파수 대역과 고주파수 대역의 에너지의 차이가 비음성구간보다 음성구간에서 크다는 관찰 결과를 이용하여 새로운 특징계수를 제안했는데, 웨이블렛 변환을 사용하여 입력신호를 두 개의 부대역으로 분할하였고, 각 부대역의 웨이블렛 계수로부터 STE를 계산하여 그 차를 SPD(Spectral Power Distance)라고 칭하였다.

$$D(i)=\left|\frac{1}{N_a}\sum_{n=1}^{N_a}X_{m,i}^2(n)-\frac{1}{N-N_a}\sum_{n=N_a+1}^{N}X_{m,i}^2(n)\right|.$$

여기서 $X_{m,i}(n)$는 웨이블렛 계수, N은 frame length, m은 scale index, i는 frame index, $N_a=N/2^m$와 $N-N_a$는 각각 저주파수 대역 및 고주파수 대역 계수의 개수이다. 참고문헌 [13]에서는 입력신호의 샘플링 주파수 $f_s=16\text{kHz}$, $m=1$을 사용하였으므로 $N_a=N/2$, $N-N_a=N/2$이고, 저대역과 고대역의 경계는 4 kHz이다.

본 논문에서는 SPD를 잡음에 더욱 강인하게 하기 위하여 각 부대역의 웨이블렛 계수의 TE를 사용하여 SPD를 구하였다.

$$D(i)=\frac{1}{N_a}\sum_{n=1}^{N_a}E_{m,i}(n)-\frac{1}{N-N_a}\sum_{n=N_a+1}^{N}E_{m,i}(n). \quad (2)$$

여기서 $E_{m,i}(n)$는 WPD(Wavelet Packet Decomposition) 계수 $X_{m,i}(n)$의 TE로서 다음과 같다.

$$E_{m,i}(n)=X_{m,i}^2(n)-X_{m,i}(n+1)X_{m,i}(n-1). \quad (3)$$

음성구간에서 특징계수 값이 매우 적을 경우를 위하여 다음과 같이 power envelope으로 증폭하였다.[13]

$$D_w(i)=D(i)\left[\frac{1}{2}+\frac{16}{\log(2)}\log\left(1+2\sum_{k=1}^{N}s_i^2(n)\right)\right]. \quad (4)$$

또, 발성이 다른 프레임들에 대한 특징계수의 영향을 균등케 하기 위하여 $D_w(i)$에 hyperbolic tangent sigmoid 함수를 적용하였다.

$$D_c(i)=\frac{1-e^{-2D_w(i)}}{1+e^{-2D_w(i)}}. \quad (5)$$

인접한 프레임 간에 특징계수의 연속성을 위하여 smoothing을 하였으며, 이에 사용된 low-pass filter

$h(i)$의 전달함수는 다음과 같다.

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

Smoothing된 특징계수 $\gamma(i)=D_c(i)*h(i)$가 본 논문에서 제안하는 특징계수이고, 이것은 SPD에 TE를 적용한 것이므로 이하 SPD-TE라고 칭한다. 본 논문에서는 $f_s=8\text{kHz}$를 사용했으므로, WPD계수들의 저대역과 고대역의 경계는 2 kHz이다. 입력 신호에 포함될 수도 있는 DC 및 저주파 잡음의 영향을 줄이기 위하여 WPD 하기 전에 차단주파수 $f_c=70$ Hz인 5th order Butterworth HPF를 사용하여 pre-emphasis를 하였다.

임계값은 주어진 프레임의 음성-비음성 여부를 판별하는 특징계수의 결정적인 값으로서, VAD에서 매우 중요한 요소이다. 정체성 잡음의 경우에는 통상적으로 처음 일정 시간 동안의 비음성 즉 잡음 구간에서 임계값을 구하여 사용한다. 그러나, 비정체성 잡음의 경우에는, 최근에 여러 가지 방안[9-14]이 제안되었지만, 특정 잡음에 대한 효용성, 또 다른 상수의 개입과 그 값의 결정 문제 등이 남는다. 여러 가지 잡음 스펙트럼의 다양한 형태와 변화 속도에 대응할 만한 적응 알고리듬은 아직 미해결 과제이다.

Pham과 Chien[13]은 statistical percentile filtering 기법을 사용한 적응 알고리듬을 제안하였다. 최근 1 s에 해당되는 프레임들의 특징계수들을 오름차 순으로 버퍼에 저장하여 변곡점에 해당되는 값을 찾아서 임계값으로 정하였다. 매 프레임마다 버퍼를 갱신하므로 임계값이 갱신된다. 그러나 이 방법이 유효하려면 변곡점이 반드시 존재해야 하고, 변곡점을 찾기 위하여 또 다른 임계값을 필요로 한다는 문제가 있다. 실제로 예비 실험 결과, SNR이 0 이하일 때에는 변곡점이 뚜렷하지 않았다. 참고문헌 [13]에도 SNR이 5dB 보다 큰 경우에 대한 실험 결과만 보였다.

Ghosh와 Narayanan[14]는 LTSV라고 하는 특징계수를 제안하였고, 임계값 갱신을 위하여 음성 버퍼의 최소값과 비음성 버퍼의 최대값의 convex sum을 사용하였다.

$$\theta(i)=\alpha\min(L_{S+N}(i))+(1-\alpha)\max(L_N(i)).$$

버퍼를 매 프레임마다 갱신하여 임계값을 갱신하였다. 초기값은 처음 1 s 동안의 비음성 프레임(100개) 특징계수들을 사용하여 $\theta=\mu_N+k\,\sigma_N$로 설정하였고, 여기서 μ_N, σ_N은 각각 그 특징계수들의 평균과 표준편차이고, k 값은 실험적으로 3을 사용하였다. 예비 실험 결과, convex sum은 본 논문에서 제안한 특징계수 $\gamma(i)$에 대하여 변별력이 약하였다. 따라서, 본 논문에서는 임계값을 다음과 같이 설정하였다.

$$\theta(i)=\mu_N(i)+k\,\sigma_N(i). \tag{7}$$

여기서 $\mu_N(i)$, $\sigma_N(i)$ 는 각각 이전 1초에 해당하는 잡음 프레임들의 $\gamma(i)$의 평균과 표준편차이다. 고정된 k 값을 사용해도 $\mu_N(i)$, $\sigma_N(i)$ 이 매 프레임마다 갱신되므로 임계값은 매 프레임마다 갱신된다. k 값을 매 프레임마다 갱신하면 adaptive threshold가 되어 비정체성 잡음에 적용할 수 있을 것이다. 본 논문에서는 정체성 잡음을 가정하여 k 값을 0부터 8까지 0.1 간격으로 가변시켜서 ROC(Receiver Operating Characteristics) 곡선으로 성능을 평가하였다.

무성음(unvoiced) 등 에너지가 작은 프레임의 판별 오류 및 고립 오류를 보정하기 위하여 hangover가 필요하다.[19,20] 본 논문에서는 ETSI 표준[20]을 사용하였다.

본 논문에서 제안한 VAD 알고리듬을 요약하면 다음과 같다. 초기 1 s 동안은 비음성 신호라고 가정하여 초기 임계값을 계산하고, 그 이후에는 1 s에 해당되는 비음성 버퍼를 갱신해가면서 식(7)에 따라 임계값을 갱신한다.

1) pre-emphasis: 70 Hz HPF.
2) window: 25 ms Hanning, 10ms advance.
3) WPD: 2 subbands, DB4(4-th order Daubechies)
4) 비음성 버퍼로부터 임계값 $\theta(i)$ 계산: 식(7).
5) 특징계수 SPD-TE 계산: 식(2)-(6) $\rightarrow\gamma(i)$.
6) $vad(i)=\begin{cases}1(\text{'speech'}), & \text{if } \gamma(i)>\theta(i)\\ 0(\text{'nonspeech'}), & \text{otherwise}\end{cases}$
7) $vad(i)=0$이면 비음성 버퍼 갱신.
8) Hangover: buffer length N=5.[20]
9) Repeat.

III. 성능 평가

음성 신호는 TIMIT 데이터베이스[21]의 core test set 중 4인의 화자(남성 2인, 여성 2인)가 녹음한 32개의 단문(texts)을 사용하였다. 단문의 길이는 2 내지 4 s 가량인데, 화자 1인당 제공된 단문 8개를 앞, 뒤, 중간에 2 s의 침묵(silence) 구간을 삽입하여 하나의 장문(long sentence)으로 길게 연결하였다. 이것은 단문 내의 짧은 구간이 아니라 긴 장문 내의 VAD를 위한 것이다. 이러한 시도는 비정체성 잡음에 대한 평가에서도 사용된 바 있다.[14] 실험에사용된 음성 신호의 속성을 Table 1에 보였다. 화자당 장문 1개의 길이는 40 s 내외, 음성의 비율은 대략 50 % 정도이다.

음성 구간의 index는 TIMIT 데이터베이스의 *.phn 파일에 있는 시작, 끝, 그리고, pause 구간의 index를 사용하였다. 잡음은 NOISEX-92 데이터베이스[22]의 잡음 중 네 가지(Babble, Factory2, Volvo, White)를 사용하였다.

입력 신호는 SNR이 10, 5, 0, -5, -10 dB가 되도록 잡음 신호를 scaling하여 음성 신호에 더한 것을 사용하였다. 음성과 잡음은 사전에 $f_s = 8$ kHz로 down-sampling 하였다.

성능 평가를 위하여 Sohn 알고리듬[7]을 비교 대상으로 선택하였는데, 그 이유는 이 알고리듬의 성능이 우수하고 계산 속도가 빨라서 최근의 거의 모든 VAD 논문들이 이 알고리듬을 비교 대상으로 선택하기 때문이다. Sohn 알고리듬에서 ROC를 구하기 위하여 음성 확률 임계값 p를 0.20부터 .99까지 0.01 간격으로 변화시켰다.

Fig. 1에는 잡음의 종류와 SNR에 대하여 네 화자의 평균 ROC 곡선을 보였다. HR1은 speech hit rate, FAR1은 speech false alarm rate이다.[10,12] 이상적으로는 HR1=1, FAR1=0이다. 그림에서 실선은 SPD-TE VAD, 점선은 Sohn's VAD이고, 각각 위로부터 SNR 10 dB부터 5 dB 간격으로 -10 dB까지 5개씩의 곡선을 보였다. Babble 잡음의 경우, 10 dB, 5 dB, 0 dB의 경우는 SPD-TE가 약간 우세하고, -5 dB, -10 dB에서는 Sohn이 우세하나, FAR1이 0.2에서 멈추는 문제점이 있다.

Table 1. Speech files used in the experiments.

Speaker	Gender	Texts	Speech Ratio (%)	Length (s)
MDAB0	M	8	47	37.745
MWBT0	M	8	53	44.730
FELC0	F	8	52	46.050
FPAS0	F	8	47	38.555

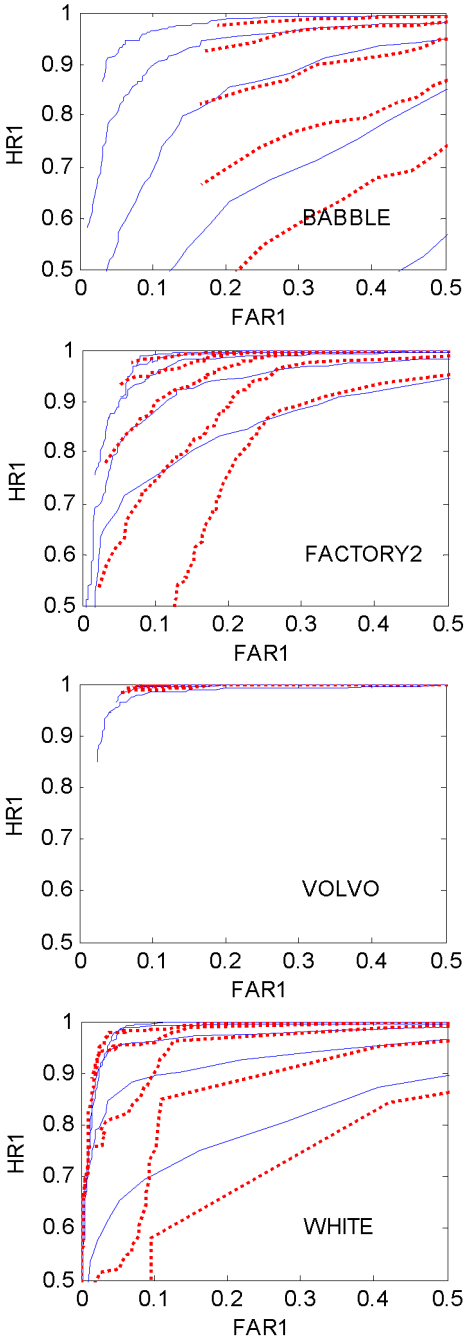
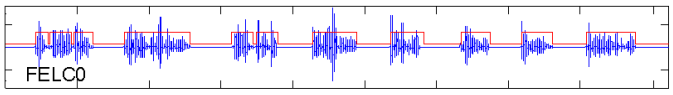


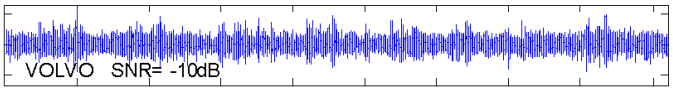
Fig. 1. Speech-averaged ROC curves for SPD-TE (solid lines) and Sohn(dotted lines) VADs for SNR =10, 5, 0, −5, −10 dB.

Table 2. Speech-averaged VAD scores.

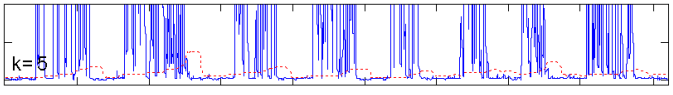
Noise	SNR	SPD-TE			Sohn		
Type	[dB]	k	FAR1	HR1	p	FAR1	HR1
Babble	10	6.6	0.0571	0.9307	0.99	0.1882	0.9763
	5	4.3	0.1081	0.9040	0.99	0.1709	0.9255
	0	3.4	0.1869	0.8314	0.99	0.1653	0.8232
	-5	2.7	0.3231	0.7126	0.98	0.2474	0.7400
	-10	1.9	0.4712	0.5243	0.96	0.3588	0.6344
Factory2	10	8.0	0.0691	0.9690	0.99	0.6700	0.9753
	5	6.8	0.0623	0.9405	0.98	0.0634	0.9416
	0	5.0	0.0636	0.9216	0.90	0.1075	0.9042
	-5	3.5	0.1140	0.8993	0.70	0.1533	0.8243
	-10	2.7	0.1888	0.8309	0.35	0.2190	0.7977
Volvo	10	8.0	0.1167	0.9995	0.99	0.0877	0.9973
	5	8.0	0.0980	0.9986	0.99	0.0773	0.9970
	0	8.0	0.0764	0.9935	0.99	0.0716	0.9961
	-5	8.0	0.0513	0.9658	0.99	0.0648	0.9925
	-10	5.0	0.0567	0.9638	0.99	0.0544	0.9825
White	10	8.0	0.0440	0.9622	0.70	0.0351	0.9695
	5	3.6	0.0446	0.9716	0.42	0.0446	0.9504
	0	2.4	0.0484	0.9521	0.30	0.0973	0.8949
	-5	2.0	0.1020	0.8950	0.23	0.1098	0.8480
	-10	1.77	0.2250	0.7670	0.23	0.2750	0.7150



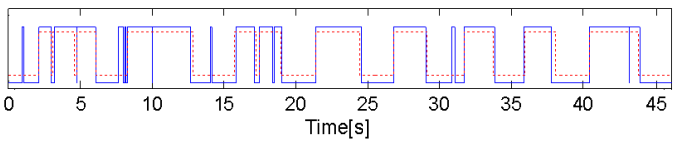
(a) Speech waveform



(b) Noisy speech with −10 dB Volvo noise



(c) SPD-TE feature values(solid lines) and threshold(dotted lines)



(d) SPD-TE(solid lines) and ideal(dotted lines) VAD results

Fig. 2. Example plots.

Factory 2와 White 잡음의 경우는 전반적으로 SPD-TE가 우세한 것으로 나타났다. Volvo 잡음의 경우에는 모두 우수하나, Sohn의 경우 FAR1이 0.1 이하로 내려가지 않았다. ROC 분석 결과, SNR= -5, -10 dB Babble 잡음을 제외한 나머지 모든 경우에 SPD-TE가 Sohn보다 우세하거나 동등한 판별 정확도를 보였다.

VAD의 최적의 동작점은 ROC 곡선과 대각선(HR1=1과 FAR1=0.5를 잇는 직선)의 교점이다. 이에 해당되는 임계값과 VAD 점수(HR1, FAR1)를 Table 2에 보였다. SNR= -5, -10 dB Babble 잡음은 Sohn이 우세하지만, Factory 2와 White 잡음의 경우는 SPD-TE가 우세하다. Volvo 잡음에 대해서는 두 가지 VAD 모두 우수한 판별 정확도를 보였다. 즉, 점수 상으로도 SPD-TE가 Sohn보다 전반적으로 동등하거나 우수한 것으로 나타났다. 이러한 점수는 이상적인 것으로써, 실제 잡음 환경에서는 적응 알고리듬이 결합되어야 얻을 수 있는 수치이다. 이는 SPD-TE와 Sohn은 물론이고 잡음의 비정체성을 전제하지 않은 모든 VAD 알고리듬에 해당된다.

VAD 동작을 보이기 위하여, Table 1의 화자 FELC0 음성 신호에 대한 -10 dB Volvo 잡음의 경우를 Fig. 2에 보였다. 이 경우에 Table 2에서 $k=5$이다. Fig. 2에서 (a)는 음성 신호이고 참고로 음성 구간을 함께 표시하였다. (b)는 잠음이 더해진 음성 신호이다. (c)는 특징계수 SPD-TE이고, 점선은 임계값이 변화하는 모습이다. 임계값이 음성 구간에서는 VAD 판별 오류의 영향으로 다소 증가하지만 비음성 구간에서는 다시 바람직하게 감소하는 모습을 볼 수 있다. (d)에서 실선은 STE-TE VAD의 최종 판별 결과이다. 점선으로 표시된 이상적인 기준에 상당히 근접함을 알 수 있다. 점수는 (FAR1, HR1)=(0.0682, 0.9483)이고, 평균 점수는 Table 2에서 (FAR1, HR1)=(0.0567, 0.9638)이다. 이 음성의 문장 길이는 46.050 s이고, SPD-TE VAD와 Sohn VAD의 run time은 Intel Core i5-2400 CPU @ 3.10 GHz에서 각각 1.266 s, 14.109 s (N=5 Hangover .064 s 포함)이다.

IV. 결 론

임계값을 이용하는 VAD 알고리듬의 핵심 요소는 특징계수와 임계값이다. 본 논문에서는 2-band WPD 계수에 Teager energy operator를 적용한 특징계수와 잡음 버퍼를 이용한 임계값 갱신 알고리듬을 사용하여 잡음에 강인한 VAD 알고리듬을 제안하였다. 음성과 잡음을 사용한 ROC 분석 결과, babble 잡음이 -5 dB 이하인 경우 외에는 대표적인 기존 알고리듬을 능가하는 판별 정확도를 보였다. 특히, Factory 2와 White 잡음의 SNR이 0 dB, -5 dB, -10 dB일 때 ROC 및 점수로 보는 판별 정확도는 현저히 개선되었다. 이렇게 낮은 SNR에서의 실험 결과를 제시한 논문이 드물다는 것을 감안하면 이는 의미있는 결과라고 할 수 있다.

VAD를 실제 비정체성 잡음 환경하에서 사용하기 위해서는 적응 알고리듬이 사용되어야 하고, 모든 경우에 적용 가능한 적응 알고리듬은 아직 미해결 과제이다. 본 논문에서 제안한 SPD-TE VAD의 경우 매 프레임마다 k 값을 갱신하는 알고리듬이 결합되면 적응알고리듬이 된다.

감사의 글

이 논문은 2011학년도 경기대학교 연구년 수혜로 연구되었음.

References

1. P. C. Loizou, *Speech Enhancement* (CRC Press, Boca Raton, 2007), pp. 309-400.
2. K. Ishizuka, T. Nakatani and N. Miyazaki, “Noise robust voice activity detection based on periodic to aperiodic component ratio,” Speech Commun.**52**, 41-60 (2010).
3. D. Ying, Y. Yan, J. Dang and F. K. Soong, “Voice activity detection based on an unsupervised learning network,” IEEE Trans. Audio, Speech, and Lang. Processing, **19**, 2624-2628 (2011).
4. T. Kristjansson, S. Deligne and P. Olsen, “Voicing features for speech detection,” in Proc. Interspeech, 369-372 (2005).
5. J-H Bach, B. Kollmeier and J. Anemuller, “Modulation-based detection of speech in real background noise: Generalization to novel background classes,” in Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. 41-44 (2010).
6. E. Chuangsuwanich and J. Glass, “Robust voice activity detector for real world application using harmonicity and modulation frequency,” in Proc. Interspeech, 2645-2648 (2011).

7. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE Signal Process. Lett. **16**, 1-3 (1999).
8. F. Beritelli, S. Casale and G. Ruggeri, "Performance evaluation and comparison of ITU-T/ETSI voice activity detectors," in Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. **3**, 1425-1428 (2001).
9. M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," IEEE Trans. Speech and Audio Process. **10**, 109-118 (2002)
10. 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).
11. 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).
12. G. Evangelopoulos and P. Maragos, "Multiband modulation energy tracking for noisy speech detection," IEEE Trans. Audio, Speech and Lang. Processing, **14**, 2024-2038 (2006).
13. 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).
14. 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).
15. James F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. **S7.3**, 381-384 (1990).
16. F. Jabloun, A. E. Cetin and E. Erzin, "Teager energy based feature parameters for speech recognition in car noises," IEEE Signal Process. Lett.. **6**, 259-261 (1999).
17. M. Bahoura and J. Rouat, "Wavelet speech enhancement based on the Teager energy operator," IEEE Signal Process. Lett. **8**, 10-12 (2001).
18. K. B. Eung, "An Experimental Study on the Robustness of the Teager Energy to the Car Noise," (in Korean), Inst. of Industrial Technology Journal, Kyonggi University, **39**, 43-56 (2011).
19. 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).
20. 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).
21. J. S. Garofolo, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium, Philadelphia, (1993).
22. 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