

잡음환경에서 Teager 에너지의 전력 스펙트럼 편차에 기반한 음성 검출 알고리즘

Voice Activity Detection Algorithm Based on the Power Spectral Deviation of Teager Energy in Noisy Environment

박 윤 식, 안 홍 섭, 이 상 민
(Yun-Sik Park, Hong-Sub An, Sangmin Lee)

인하대학교 전자공학부

(접수일자: 2011년 7월 8일; 수정일자: 2011년 8월 4일; 채택일자: 2011년 9월 8일)

본 논문에서는 잡음환경에서 효과적인 음성을 검출하기 위한 새로운 음성검출 (VAD, voice activity detection) 알고리즘을 제안한다. 제안된 방법은 개선된 음성/비음성 검출을 위해 기존의 파워 스펙트럼 편차를 적용하는 대신 Teager 에너지 기반의 파워 스펙트럼 편차 (power spectral deviation)를 이용한다. 또한 향상된 VAD 성능을 위하여 각각의 주파수 밴드에 대한 음성부재확률 (speech absence probability)을 제안된 파워 스펙트럼 편차를 도출하는데 스무딩 (smoothing) 파라미터로 적용한다. 제안된 알고리즘은 기존의 방법과 객관적인 실험을 통해 비교 평가한 결과 다양한 배경잡음 환경에서 향상된 성능을 보였다.

핵심용어: 전력 스펙트럼 편차, Teager 에너지, 음성부재확률

투고분야: 음성처리 분야 (2,3)

In this paper, we propose a novel voice activity detection (VAD) algorithm to effectively distinguish speech from nonspeech in various noisy environments. The presented VAD utilizes the power spectral deviation (PSD) based on Teager energy (TE) instead of the conventional PSD scheme to improve the performance of decision for speech segments. In addition, the speech absence probability (SAP) is derived in each frequency subband to modify the PSD for further VAD. Performances of the proposed VAD algorithm are evaluated by objective test under various environments and better results compared with the conventional methods are obtained.

Keywords: Power spectral deviation, Teager energy, Speech absence probability

ASK subject classification: Speech Signal Processing (2,3)

I. 서론

음성검출 (VAD, voice activity detection) 알고리즘은 음성인식이나 잡음제거기 등 다양한 음성처리 시스템에 적용되고 있으며 시스템 성능에 주요한 영향을 미치는 핵심 부분으로 인식되고 있다 [1]. 특히 직접적으로 VAD를 통해 잡음신호를 추정하여 잡음을 제거하는 잡음제거 (NS, noise suppression) 알고리즘에서는 VAD 성능이 NS 알고리즘 성능에 전반적으로 영향을 주는 주요한 요소라 할 수 있다. 일반적으로 VAD 알고리즘은 음성/비음성 신호를 판별 할 수 있는 피쳐 (feature)를 도출하고

적절한 문턱 (theshold) 값을 피쳐에 적용하는 결정식 (decision rule)의 형태로 음성과 비음성을 구분한다. 이러한 VAD에 사용되는 피쳐로는 비교적 보편적으로 사용되는 스펙트럼 에너지 (spectral energy), 전력 스펙트럼 편차 (PSD, power spectral deviation), ZCR (zero-crossing ratio)에서부터 LPC (linear prediction coefficients) 및 통계적 모델에 기반한 likelihood ratio (LR)등 다양한 피쳐들이 적용되고 있다 [2]. 하지만 VAD를 위해 적용되는 피쳐들은 신호 대 잡음 비 (SNR, signal-to-noise ratio)가 큰 환경에서는 음성/비음성에 대한 피쳐 특성이 비교적 분명하지만 다양한 배경 잡음이 존재하는 실제 잡음환경이나 SNR이 낮은 신호에 대해서는 피쳐들이 잡음신호에 민감하기 때문에 피쳐에 의한 VAD의 성능이 저하되

는 문제점이 발생한다. 따라서 본 논문에서는 SNR이 낮은 잡음환경에서도 효과적인 음성검출 성능을 보이는 Teager energy (TE)를 적용한 PSD 기반의 VAD 알고리즘을 제안한다. 구체적으로 제안된 방법에서는 잡음을 제거하는 효과로 인해 잡음에 대한 음성신호의 특성을 강화시킴으로서 음성검출의 성능을 향상시킬 수 있는 TE operator를 시간영역에서 적용하고 도출된 TE에 대한 주파수 영역에서의 PSD를 VAD를 위한 피쳐로 사용한다. 또한 잡음환경에 강인한 PSD를 위하여 TE 기반의 음성부재확률 (SAP, speech absence probability)을 PSD 도출을 위한 스무딩 (smoothing) 파라미터로 적용한다. 제안된 방법의 성능 평가를 위해 음성 스펙트로그램 (spectrogram)과 ITU-T P.826 perceptual evaluation of speech quality (PESQ) 및 composite measure 테스트를 실시하였고 제안된 기법은 다양한 잡음 환경에서 기존의 방법보다 우수한 성능을 보였다.

II. Teager Energy Operator

이 장에서는 제안된 VAD에서 TE 기반의 PSD를 도출하기 위한 TE operator에 대해 간략하게 설명한다. TE operator는 다양한 잡음 환경에서 오염된 음성신호에 대하여 잡음신호는 감쇄시키고 음성신호는 증폭시켜주는 효과를 보임으로서 잡음신호에 대한 음성신호의 특성을 강화시켜 주는 연사자이다. 따라서 잡음환경에서 보다 효과적인 음성검출을 위해 잡음과 음성신호의 특성 차이를 보일 수 있는 다양한 피쳐 파라미터들을 추정하는데 널리 적용되어 왔다 [3-4]. 구체적으로 continuous 시간에서의 신호를 $s(t)$ 라고 한다면 TE operator는 다음과 같이 정의된다.

$$\Psi_c[s(t)] = [\dot{s}(t)]^2 - s(t)\ddot{s}(t) \quad (1)$$

여기서 $\dot{s} = ds/dt$ 이며 discrete 시간에서의 TE operator는 다음과 같이 표현된다 [3-4].

$$\Psi_d[s(n)] = s(n)^2 - s(n+1)s(n-1) \quad (2)$$

여기서 n 은 discrete 시간에서의 시간 index를 의미한다. 실제 잡음환경을 고려하여 배경잡음에 의해 오염된 마이크로폰 입력신호 $y(n)$ 은 다음과 같이 나타낼 수 있다.

$$y(n) = s(n) + d(n) \quad (3)$$

여기서 $s(n)$ 과 $d(n)$ 은 각각 깨끗한 음성신호와 부가된

잡음신호를 의미하며 $s(n)$ 와 $d(n)$ 은 상관관계가 없다고 가정하면 오염된 입력신호 $y(n)$ 의 TE $\Psi_d[y(n)]$ 은 다음과 같이 나타 낼 수 있다.

$$\Psi_d[y(n)] = \Psi_d[s(n)] + \Psi_d[d(n)] + 2\tilde{\Psi}_d[s(n), d(n)] \quad (4)$$

여기서 $\Psi_d[s(n)]$ 와 $\Psi_d[d(n)]$ 은 각각 깨끗한 음성신호와 부가된 잡음신호에 대한 TE를 의미하며 $s(n)$ 과 $d(n)$ 의 cross- Ψ_d 에너지 $\tilde{\Psi}_d[s(n), d(n)]$ 는 다음과 도출된다.

$$\tilde{\Psi}_d[s(n), d(n)] = s(n)d(n) - 0.5s(n-1)d(n+1) - 0.5s(n+1)d(n-1) \quad (5)$$

여기서 $s(n)$ 과 $d(n)$ 은 zero mean이고 서로 독립이라 가정 하면 $\tilde{\Psi}_d[s(n), d(n)]$ 의 기대 값은 0이 되고, 따라서 식 (4)는 다음과 같이 간략화 된다.

$$E\{\Psi_d[y(n)]\} = E\{\Psi_d[s(n)]\} + E\{\Psi_d[d(n)]\} \quad (6)$$

여기서 음성신호의 TE는 잡음신호의 TE보다 값이 상당히 크기 때문에 최종적으로 식 (6)은 다음과 같이 나타 낼 수 있다.

$$E\{\Psi_d[y(n)]\} \approx E\{\Psi_d[s(n)]\} \quad (7)$$

TE의 이러한 특성 때문에 TE operator를 거친 오염된 음성신호는 잡음신호가 제거된 효과를 얻을 수 있으며 따라서 기존의 보통 (regular) 에너지에 기반한 피쳐파라미터 보다 TE 기반의 피쳐 파라미터들이 잡음환경에서 보다 나은 음성 특성을 도출할 수 있다.

III. 제안된 Teager Energy를 이용한 파워 스펙트럼 편차 기반의 VAD 알고리즘

2장에서는 시간 영역에서 TE operator에 대하여 간략히 설명하였다. 3장에서는 TE를 이용하여 주파수영역에서 PSD기반의 새로운 VAD 알고리즘을 제안한다. 또한 제안된 알고리즘에서는 잡음환경에서 보다 효과적인 PSD를 위하여 TE 기반의 SAP를 PSD에 스무딩 파라미터로 적용한다. 일반적으로 PSD는 음성/비음성의 특성을 비교적 잘 나타내는 피쳐로서 다양한 VAD 알고리즘에 널리 사용되어 왔다. 하지만 PSD 또한 음성신호에 비해 잡음신호가 강한 낮은 SNR 환경에서는 잡음신호의 영향을 받아 잡음신호에 대한 음성신호의 특성이 약해지는 단점이 발생한다. 따라서 본 논문에서 잡음환경에서 효과적인 PSD를 구하기 위해 시간 영역에서 TE operator에 의

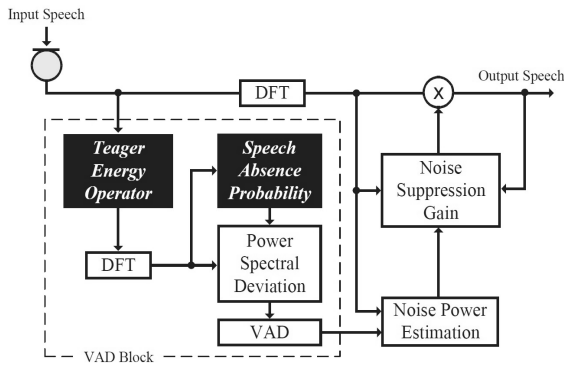


그림 1. 제안된 VAD 알고리즘의 블록도
Fig. 1. Block diagram of the proposed VAD algorithm.

해 잡음에 대한 음성신호의 특성이 분명해진 입력신호에 대하여 새로운 TE 기반의 PSD를 도출하여 VAD를 위한 피쳐 파라미터로 적용한다. 그림 1은 제안된 VAD 알고리즘의 블록도를 보여주고 있으며 그림 1로부터 잡음전력 추정을 위한 VAD 알고리즘에서 입력신호는 DFT (discrete Fourier transform) 전에 TE operator를 거치고 주파수 도메인에서 SAP가 도출되어 제안된 PSD를 구하는데 적용되는 것을 볼 수 있다. 구체적으로, 제안된 TE 기반의 PSD $\Delta_{TE}(i)$ 는 다음과 같이 도출된다 [5].

$$\Delta_{TE}(i) = \sum_{k=0}^{M-1} |Y_{dB}(i, k) - \bar{Y}_{dB}(i, k)| \quad (8)$$

여기서 i 와 k 는 각각 128-point DFT를 거친 주파수 영역에서 프레임과 주파수 index를 나타내며 $M(=64)$ 는 각 프레임에서 주파수 index의 전체 개수를 의미한다. 또한 $\bar{Y}_{dB}(i, k)$ 는 TE 기반의 로그 전력 스펙트럼으로서 다음과 같이 추정된다.

$$Y_{dB}(i, k) = 10 \log_{10} (|\Psi[Y(i, k)]|^2) \quad (9)$$

여기서 $\Psi[Y(i, k)]$ 는 TE operator를 거친 입력신호의 DFT 스펙트럼을 나타낸다. 식 (8)에서 $\bar{Y}_{dB}(i, k)$ 는 이전 프레임의 주파수 밴드에서 추정된 long-term averaging 전력 스펙트럼으로 본 논문에서 잡음환경에서 보다 효과적인 PSD를 도출하기 위해 다음과 같이 TE에 기반한 SAP $p(H_0|\Psi[Y(i, k)])$ 를 스무딩 파라미터로 적용한다.

$$\bar{Y}_{dB}(i, k) = (1 - p(H_0|\Psi[Y(i, k)])) \bar{Y}_{dB}(i-1, k) + p(H_0|\Psi[Y(i, k)]) Y_{dB}(i, k) \quad (10)$$

여기서 TE기반의 SAP를 도출하기 위하여 TE 스펙트럼 $\Psi[Y(i, k)]$ 에 대하여 깨끗한 음성신호와 잡음신호가 상관관계가 없고 신호의 통계적 모델이 일반적으로 통계적

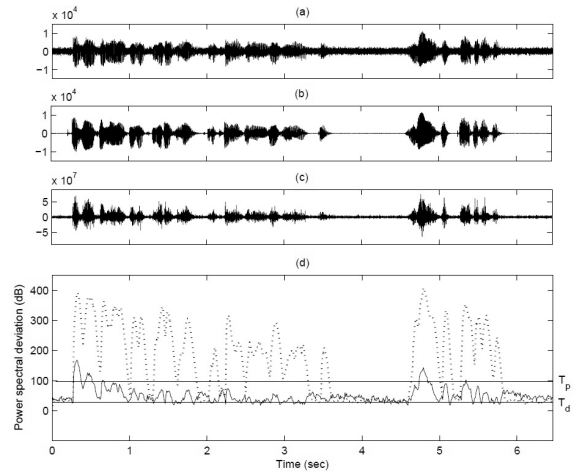


그림 2. 전력 스펙트럼 편차의 비교 (white noise, SNR=5 dB)
(a) 오염된 음성신호 파형 (b) 깨끗한 음성신호 파형
(c) Teager energy 파형 (d) 전력 스펙트럼 편차: 기존의 방법 (실선), 제안된 방법 (점선)

Fig. 2. Comparison of power spectral deviation (white noise, SNR=5 dB) (a) Noisy speech waveform (b) Clean speech waveform (c) Teager energy waveform (d) Power spectral deviation: the conventional method (solid line), the proposed method (dashed line).

모델 기반의 파라미터 추정 방법에서 널리 적용되고 있는 zero-mean complex Gaussian 분포로 특징지어진다는 가정 하에 음성신호가 존재하지 않을 때와 존재할 경우 각각의 조건 H_0, H_1 에서 $p(H_0|\Psi[Y(i, k)])$ 는 Bayes' rule을 적용하여 다음과 같이 도출된다 [6-8].

$$p(H_0|\Psi[Y(i, k)]) = \frac{p(\Psi[Y(i, k)]|H_0)p(H_0)}{p(\Psi[Y(i, k)]|H_0)p(H_0) + p(\Psi[Y(i, k)]|H_1)p(H_1)} \quad (11)$$

$$= \frac{1}{1 + \frac{p(H_1)}{p(H_0)} \Lambda(\Psi[Y(i, k)])}$$

여기서 $p(H_0)(=1-p(H_1))$ 는 음성부재에 대한 a priori probability를 나타내고 $\Lambda(\Psi[Y(i, k)])$ 는 likelihood ratio 로서 다음과 같이 나타낸다.

$$\Lambda(\Psi[Y(i, k)]) = \frac{p(\Psi[Y(i, k)]|H_1)}{p(\Psi[Y(i, k)]|H_0)} \quad (12)$$

최종적으로 본 논문에서 제안하는 VAD 알고리즘은 TE 기반 PSD (TE-PSD)에 문턱 (threshold) 값을 적용한 결정식에 의해서 음성/비음성 신호를 판별할 수 있도록 다음과 같이 표현된다.

$$f_{VAD} = \begin{cases} \text{음성 (speech)}, & \text{if } \Delta_{TE}(i) > T \\ \text{비음성 (non-speech)}, & \text{otherwise} \end{cases} \quad (13)$$

그림 2의 (c)는 마이크론의 오염된 입력신호에 대하여 TE operator를 적용한 파형을 보여주고 있으며 그림

2의 (c)로부터 TE는 오염된 입력신호보다 상대적으로 잡음은 상당히 제거되고 음성신호는 잡음신호에 비해 특성이 뚜렷해지는 것을 볼 수 있다. 또한 그림 2의 (d)는 제안된 식 (13)의 결정 식에 의해 제안된 TE-PSD에 문턱 값 $T_p (=83)$ 을 적용한 것을 보여주는 것으로서 상용코덱인 IS-127 EVRC (enhanced variable rate codec)에 포함된 잡음제거기에서 VAD 피쳐로 사용되는 기존의 PSD에 문턱 값 $T_d (=28)$ 를 적용한 것보다 개선된 PSD 특성을 나타내는 것을 알 수 있다. 그리고 이러한 VAD 알고리즘을 기반으로 잡음제거기에서 VAD 결과가 해당 구간을 비음성으로 결정하였을 경우 averaging rule을 통해 비음성 구간에서 잡음전력 $\lambda_d(i, k)$ 은 다음과 같이 업데이트되어 추정된다.

$$\lambda_d(i, k) = \alpha_d \lambda_d(i-1, k) + (1 - \alpha_d) |Y(i, k)|^2 \quad (14)$$

여기서 α_d 는 스무딩 파라미터로 0.99로 설정되고, 성능평가를 위하여 제안된 VAD 알고리즘은 다음과 같이 minimum mean square error (MMSE) 기반의 잡음제거 이득 $G(\xi(i, k), \gamma(i, k))$ 을 가지는 잡음제거기에 적용된다 [9].

$$\hat{S}(i, k) = G(\xi(i, k), \gamma(i, k)) Y(i, k) \quad (15)$$

여기서 $\hat{S}(i, k)$ 는 추정된 깨끗한 음성신호를 의미하며 잡음제거 이득 $G(\xi(i, k), \gamma(i, k))$ 는 다음과 같이 도출된다.

$$G(\xi(i, k), \gamma(i, k)) = \frac{\sqrt{\pi v(i, k)}}{2\gamma(i, k)} \exp\left(-\frac{v(i, k)}{2}\right) \left[(1 + v(i, k)) I_0\left(\frac{v(i, k)}{2}\right) + v(i, k) I_1\left(\frac{v(i, k)}{2}\right) \right] \quad (16)$$

식 (16)에서 I_0 와 I_1 는 각각 0차와 1차의 modified Bessel 함수를 의미하며 $v(i, k)$ 는 다음과 같다.

$$v(i, k) = \frac{\xi(i, k)}{1 + \xi(i, k)} \gamma(i, k) \quad (17)$$

여기서 $\xi(i, k)$ 와 $\gamma(i, k)$ 는 각각 *a priori* SNR과 *a posteriori* SNR을 의미하며 다음과 같이 정의된다.

$$\gamma(i, k) \equiv \frac{|Y(i, k)|^2}{\lambda_d(i, k)} \quad (18)$$

$$\xi(i, k) \equiv \frac{\lambda_s(i, k)}{\lambda_d(i, k)} \quad (19)$$

식 (19)에서 $\lambda_s(i, k)$ 는 decision-directed 기법에 의해 추정된 깨끗한 음성의 전력을 나타낸다.

IV. 실험 및 결과고찰

본 논문에서는 제안된 알고리즘의 성능 평가를 위해 다양한 잡음환경에서 객관적인 실험을 수행하였다. 성능평가는 스펙트럼 분석을 위한 음성 스펙트로그램 (spectrogram)와 객관적인 음질평가인 PESQ (perceptual evaluation of speech quality) [10] 및 다음과 같이 전반적인 음질 C_{ovl} 로 표현되는 composite measure 테스트를 실시하였다 [11].

$$C_{ovl} = 1.594 + 0.805 S_{PESQ} - 0.512 S_{LLR} - 0.007 S_{WSS} \quad (20)$$

여기서 S_{PESQ} 는 PESQ를 의미하며, S_{LLR} , S_{WSS} 는 각각 log-likelihood ratio (LLR)과 weighted-slope spectral (WSS)를 나타낸다. 테스트 샘플을 위해 8명의 화자로부터 얻은 8 kHz로 샘플링 된 60개의 깨끗한 음성신호를 수집하고 이를 세가지형태의 잡음을 다양한 SNR로 부가하여 오염된 음성신호를 생성하였으며 부가된 잡음은 NOISEX-92 데이터베이스의 white, babble, vehicle 잡음으로 SNR은 0, 5, 10, 15 dB로 달리하였다. 또한 주파수 영역에서 64개의 주파수 index는 계산량의 효율성을 고려하여 IS-127의 잡음제거기에서 narrow 밴드의 음성대역 (4 kHz)에 따른 16개의 주파수 서브밴드 (subband)로 적용하였다.

제안된 TE-PSD 기반 VAD의 성능평가는 제안된 알고리즘을 MMSE 기반의 잡음제거 이득을 사용하는 잡음제거기에 적용하고 도출된 결과 음성신호에 대하여 평가하였으며 성능 비교를 위하여 IS-127의 NS 알고리즘에서 VAD 피쳐 중 하나로 사용되는 기존의 PSD와 잡음환경에서 비교적 우수한 성능을 보인다는 통계적 모델을 이용한 likelihood ratio (LR) 기반의 VAD 알고리즘을 동일한 잡음제거기에 각각 적용하여 잡음이 제거된 결과신호에 대한 음성 스펙트로그램 및 객관적인 음질 평가 수치로부터 각각의 VAD의 성능을 비교평가 위한 결과를 도출하였다 [5, 12]. 그림 3과 그림 4는 각각의 알고리즘에 대한 음성 스펙트로그램을 보여주고 있다. 특히, 그림 3의 (e)와 그림 4의 (e)는 제안된 VAD 알고리즘이 적용된 잡음제거기로부터 도출된 결과신호에 대한 음성 스펙트로그램으로 기존의 방법이 적용된 결과신호의 음성 스펙트로그램보다 향상된 잡음제거 성능을 보이는 것을 볼 수 있다. 또한 각각의 알고리즘에 의한 PESQ 수치를 보여주고 있는 표 1로부터 모든 잡음환경에 대하여 제안된 알고리즘이 기존의 알고리즘 보다 향상된 PESQ 수치를 보여 주고 있는 것을 알 수 있다. 최종적으로 표 2는 composite measure

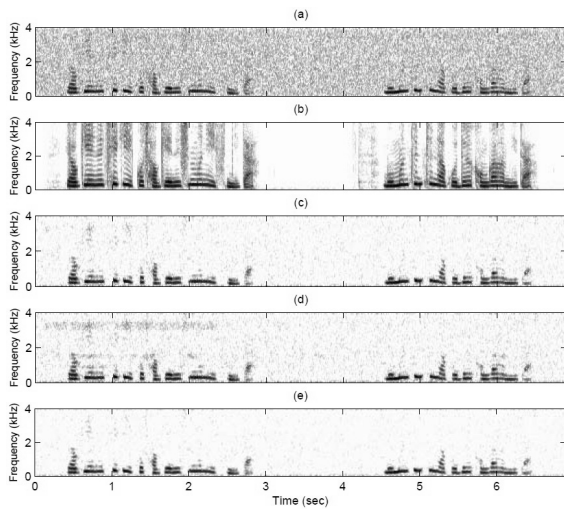


그림 3. 음성 스펙트로그램 (white noise, SNR=5 dB) (a) 오염된 음성신호 (b) 깨끗한 음성신호 (c) LR 기반의 결과신호 (d) 기존의 PSD 기반의 결과신호 (e) 제안된 TE-PSD 기반의 결과신호

Fig. 3. Speech spectrograms (white noise, SNR=5 dB) (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the method based on LR (d) Output signal obtained by the method based on PSD (e) Output signal obtained by TE-PSD.

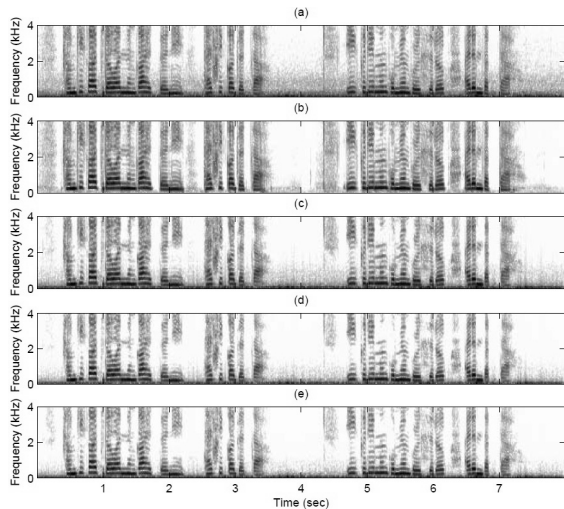


그림 4. 음성 스펙트로그램 (vehicle noise, SNR=5 dB) (a) 오염된 음성신호 (b) 깨끗한 음성신호 (c) LR 기반의 결과신호 (d) 기존의 PSD 기반의 결과신호 (e) 제안된 TE-PSD 기반의 결과신호

Fig. 3. Speech spectrograms (vehicle noise, SNR=5 dB) (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the method based on LR (d) Output signal obtained by the method based on PSD (e) Output signal obtained by TE-PSD.

에 대해 수치를 보여주고 있다. 표 2로부터 제안된 알고리즘이 기존의 방법보다 모든 잡음 환경에서 개선된 composite measure 결과를 보여 주고 있는 것을 알 수 있다. 따라서 잡음신호가 음성신호에 비해 비교적 작은

표 1. 다양한 잡음환경에서 기존과 제안된 알고리즘에 대한 PESQ 수치

Table 1. PESQ scores of the conventional method and the proposed method under noise environments.

| Noise type | SNR (dB) | Method | | |
|------------|----------|--------|-------|--------|
| | | LR | PSD | TE-PSD |
| White | 0 | 1.657 | 1.658 | 1.659 |
| | 5 | 2.083 | 2.073 | 2.088 |
| | 10 | 2.432 | 2.432 | 2.437 |
| | 15 | 2.757 | 2.753 | 2.760 |
| Babble | 0 | 1.908 | 1.888 | 1.913 |
| | 5 | 2.301 | 2.300 | 2.318 |
| | 10 | 2.641 | 2.649 | 2.659 |
| | 15 | 2.946 | 2.958 | 2.962 |
| Vehicle | 0 | 3.054 | 3.048 | 3.094 |
| | 5 | 3.366 | 3.357 | 3.401 |
| | 10 | 3.650 | 3.638 | 3.673 |
| | 15 | 3.894 | 3.875 | 3.908 |

표 2. 다양한 잡음환경에서 기존과 제안된 알고리즘에 대한 composite measure 수치

Table 2. Composite measure scores of the conventional method and the proposed method under noise environments.

| Noise type | SNR (dB) | Method | | |
|------------|----------|--------|-------|--------|
| | | LR | PSD | TE-PSD |
| White | 0 | 1.797 | 1.790 | 1.799 |
| | 5 | 2.305 | 2.284 | 2.312 |
| | 10 | 2.714 | 2.713 | 2.721 |
| | 15 | 3.078 | 3.073 | 3.083 |
| Babble | 0 | 2.173 | 2.138 | 2.197 |
| | 5 | 2.638 | 2.622 | 2.662 |
| | 10 | 3.029 | 3.037 | 3.053 |
| | 15 | 3.371 | 3.389 | 3.392 |
| Vehicle | 0 | 3.370 | 3.360 | 3.422 |
| | 5 | 3.726 | 3.713 | 3.773 |
| | 10 | 4.045 | 4.027 | 4.080 |
| | 15 | 4.318 | 4.291 | 4.340 |

SNR=15 dB의 잡음환경에서부터 잡음신호의 의해 상당히 오염된 환경인 SNR=0 dB까지의 다양한 잡음환경에서 모든 잡음 종류에 대해 표 1과 표 2에서 제안된 알고리즘이 기존의 방법보다 향상된 음질평가 수치를 보임으로서 제안된 VAD 알고리즘이 잡음제거기의 성능을 향상시킨 것을 알 수 있다.

V. 결론

본 논문에서는 잡음환경에서 효과적인 음성/비음성 검출을 위해 TE를 이용한 PSD 기반의 새로운 VAD 알고리

음을 제안하였다. 또한 제안된 방법에서는 효과적인 TF-PSD를 도출하기 위해 SAP를 스무딩 파라미터로 적용하였으며 객관적인 테스트로부터 제안된 알고리즘이 기존의 방법보다 향상된 결과를 나타내었다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 IT 융합 고급인력과정 지원사업의 연구결과로 수행되었음 (NIPA-2011-C6150-1102-0001).

참고 문헌

1. L. Karray, C. Mokbel and J. Monne, "Solutions for robust, speech/non-speech detection in wireless environment," presented at the IVTTA, 1988.
2. L. R. Rabiner and M. R. Sambur, "Voiced-unvoiced-silence detection using the Itakura LPC distance measure," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 323-326, 1977.
3. F. Jabloun, A. E. Cetin and E. Erzin, "Teager energy based feature parameters for speech recognition in car noise," *IEEE Signal Processing Letters*, vol. 6, pp. 259-261, 1999.
4. K. C. Wang and Y. H. Tsai, "Voice activity detection algorithm with low signal-to-noise ratios based on spectrum entropy," *Second International Symposium on Universal Communication 2008*, pp. 423-428, 2008.
5. TIA/EIA/IS-127, Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, 1996.
6. R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 137-145, 1980.
7. J. Sohn, W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 365-368, 1998.
8. 박윤식, 장준혁, "강인한 음성향상을 위한 Minimum Statistics와 Soft Decision의 확률적 결합의 새로운 잡음전력 추정기법," *한국음향학회지*, 26권, 4호, 153-158쪽, 2007.
9. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109-1121, 1984.

10. Rix, A. W., Beerends, J. G., Hollier, M. P. and Hekstra, A. P. "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2, pp. 749-752, 2001.
11. Yi Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. ASLP*, vol. 16, pp. 229-238, 2008.
12. J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1-3, 1999.

저자 약력

•박 윤 식 (Yun-Sik Park)



2006년 2월: 인하대학교 전자공학과 학사
 2008년 2월: 인하대학교 전자공학부 석사
 2008년 3월 ~ 현재: 인하대학교 전자공학부 박사 과정

•안 홍 섭 (Hong-Sub An)



2010년 2월: 인하대학교 전자공학과 학사
 2010년 3월 ~ 현재: 인하대학교 전자공학부 석사 과정

•이 상 민 (Sangmin Lee)



1987년: 인하대학교 전자공학과 학사 졸업
 1989년: 인하대학교 전자공학과 석사 졸업
 2000년: 인하대학교 전자공학과 박사 졸업
 1989년 1월 ~ 1994년 7월: LG 이노텍 선임연구원
 1995년 1월 ~ 2002년 3월: 삼성종합기술원 책임연구원
 2002년 4월 ~ 2005년 2월: 한양대학교 의공학교실 연구교수
 2005년 3월 ~ 2006년 8월: 전북대학교 생체정보공학부 조교수
 2006년 9월 ~ 현재: 인하대학교 전자전기공학부 부교수