논문 2012-49SP-3-10

잡음환경에서 Teager 에너지와 음성부재확률 기반의 음성향상 알고리즘

(Speech Enhancement Algorithm Based on Teager Energy and Speech Absence Probability in Noisy Environments)

박 윤 식*, 안 홍 섭*, 이 상 민**

(Yun-Sik Park, Hong-Sub An and Sangmin Lee)

요 약

본 논문에서는 다양한 잡음환경에서 효과적인 잡음 제거 (NS, noise suppression)를 위한 새로운 음성향상 (speech enhancement) 알고리즘을 제안한다. 제안된 방법에서는 음성향상 알고리즘에서 잡음전력 갱신을 위한 음성검출 (VAD, voice activity detection)의 피쳐 (feature) 파라미터로서 오염된 음성신호를 기반으로 주파수 밴드 별로 도출되는 기존의 지역 음성부재확률 (LSAP, local speech absecne probability) 대신 오염된 음성신호의 Teager energy (TE)를 적용한 LSAP를 적용한다. 또한 적용된 TE operator의 성능을 개선하기 위하여 프레임 단위로 도출되는 전역 음성부재확률 (GSAP, global SAP)을 TE의 가중치 파라미터로서 적용한다. 제안된 알고리즘은 기존의 방법과 객관적인 실험을 통해 비교 평가한 결과 다양한 배경잡음 환경에서 향상된 성능을 보였다.

Abstract

In this paper, we propose a novel speech enhancement algorithm for effective noise suppression in various noisy environments. In the proposed method, to result in improved decision performance for speech and noise segments, local speech absence probability (LSAP, local SAP) based on Teager energy of noisy speech is used as the feature parameter for voice activity detection (VAD) in each frequency subband instead of conventional LSAP. In addition, The presented method utilizes global SAP (GSAP) derived in each frame as the weighting parameter for the modification of the adopted TE operator to improve the performance of TE operator. Performances of the proposed algorithm are evaluated by objective test under various environments and better results compared with the conventional methods are obtained.

Keywords : 음성향상, 음성부재확률, Teager energy, 음성검출

Ⅰ. 서 론

음성향상 (speech enhancement) 알고리즘은 배경잡음이 존재하는 잡음환경에서 음성부호화기나 음성인식기와 같은 다양한 음성처리 시스템의 전처리기로 사용되어 시스템의 성능개선이나 직접적인 통화품질의 향상시킴으로서 주요한 부분으로 인식되고 있다 [1]. 따라서 음성향상 알고리즘의 성능향상을 위해 통계적 모델 적용 [2], 잡음제거 이득에 대한 수정 및 잡음전력 추정 등을 포함하여 여러 관련 기법이 연구되어 왔으며 특히, 추정된 잡음전력은 음성향상 알고리즘의의 다양한 파라미터에 적용되기 때문에 음성향상 알고리즘 성능에 전반적으로 영향을 주는 주요한 파라미터로서 잡음전력 추정에 대한 다양한 기법들이 제시되었다. 일반적으로

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Department of Electronic Engineering, Inha University)

※ 본 연구는 지식경제부 바이오 의료기기 전략기술개발사업의 지원(과제번호: 10031764)에 의하여 이루어졌음.

접수일자: 2011년8월9일, 수정완료일: 2012년2월9일

잡음전력은 음성검출 (VAD, voice activity detection) 알고리즘에 의하여 잡음 구간에서만 잡음신호에 대한 갱신을 통하여 추정되기 때문에 잡음전력 추정은 VAD 알고리즘의 성능에 크게 영향을 받는다고 할 수 있다[3].

구체적으로 VAD 알고리즘은 잡음신호로부터 음성신호를 판별 할 수 있는 피쳐 (feature) 파라미터를 도출하고 적절한 문턱 (theshold) 값을 피쳐 파라미터에 적용하여 문턱 값을 기준으로 음성과 비음성이 결정되는 결정식 (decision rule)으로 구성된다. 이러한 VAD에 사용되는 피쳐 파라미터로는 비교적 보편적으로 사용되는 스펙트럼 에너지 (spectral energy) 또는 ZCR (zero-crossing ratio)에서부터 LPC (linear prediction coefficients) 및 통계적 모델에 기반한 likelihodd ratio (LR)등 다양한 피쳐 파라미터들이 적용되고 있다. 하지만 VAD를 위해 적용되는 피쳐 파라미터들은 신호 대 잡음 비 (SNR, signal-to-noise ratio)가 큰 환경에서는 잡음으로부터 음성에 대한 피쳐 파라미터의 특성이 비교적 분명하지만 다양한 배경 잡음이 존재하는 실제 잡음환경이나 SNR이 낮은 음성신호에 대해서는 피쳐 파라미터들이 잡음신호에 민감하기 때문에 VAD의 성능이 저하되는 문제점이 발생한다[4].

따라서 본 논문에서는 다양한 잡음환경에서 효과적인 음성검출을 위하여 Teager energy (TE)와 음성부재확률 (SAP, speech absence probability)을 기반으로 도출되는 피쳐 파라미터를 적용한 VAD 알고리즘을 제안한다. 구체적으로 제안된 방법에서는 잡음신호에 의해 오염된 음성신호를 그대로 적용하는 기존의 방법대신 잡음을 제거하여 잡음에 대한 음성의 특성을 강화시킬 수 있는 TE operator[4~5]를 시간영역에서 적용하고 강화된 입력신호의 TE에 대한 주파수 영역에서의 통계적 모델 기반의 likelihood ratio (LR)구하고 LR로부터 주파수 밴드별로 도출되는 지역 음성부재확률 (LSAP, local SAP)을 VAD를 위한 피쳐 파라미터로 사용한다[6]. 또한 SNR 낮은 구간에서 기존의 TE operator에 의해 잡음과 함께 음성이 심하게 제거되어 발생될 수 있는 음성 왜곡 현상을 개선하기 위해 LR로부터 프레임 단위로 도출되는 전역 음성부재확률 (GSAP, global SAP)[7~10]을 TE를 구하기 위한 가중치 파라미터로 적용한다. 제안된 방법의 성능은 음성 스펙트로그램 (spectrogram)과 ITU-T P.826 perceptual evaluation of speech quality (PESQ) 및 composite measure[11]에 의해 평가되었으며 다양한 잡음 환경에서 기존의 방법보다 우수한 성능을 보였다.

Ⅱ. Teaser Energy Operator

Ⅱ장에서는 제안된 음성향상 알고리즘에서 TE 도출을 위한 TE operator에 대해 간략하게 설명한다. TE operator는 잡음을 제거하여 잡음신호에 대한 음성신호의 특성을 강화시킴으로서 잡음환경에서 보다 강인한 피쳐 파라미터들을 도출하기 위해 널리 적용되는 알고리즘이다[4~5]. 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[4~5].

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

여기서 n은 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 $\Psi[y(n)]$은 다음과 같이 나타 낼 수 있다.

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

여기서 $\Psi[s(n)]$와 $\Psi[d(n)]$은 각각 깨끗한 음성신호와 부가된 잡음신호에 대한 TE를 의미하며 $s(n)$과 $d(n)$의 상관 에너지 $\tilde{\Psi}[s(n), d(n)] = s(n)d(n) - 0.5s(n-1)d(n+1) - 0.5s(n+1)d(n-1)$이다.

$s(n)$과 $d(n)$은 zero mean이며 서로 독립이라고 가정하면 $\tilde{\Psi}[s(n), d(n)]$의 기대 값은 0이 되고, 따라서 식 (4)는 다음과 같이 표현된다.

$$E\{\Psi[y(n)]\} = E\{\Psi[s(n)]\} + E\{\Psi[d(n)]\} \quad (5)$$

여기서 음성신호의 TE $\Psi[s(n)]$는 잡음신호의 TE

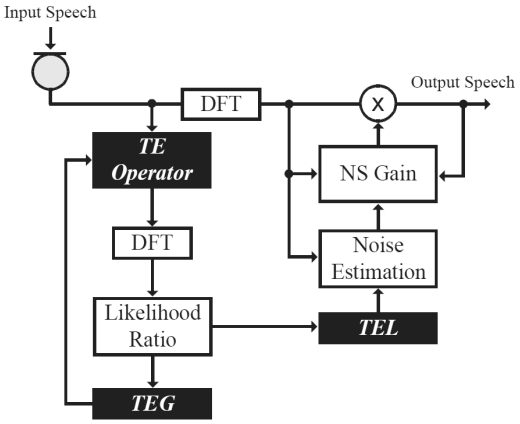

그림 1. 제안된 음성향상 알고리즘의 블록도.
Fig. 1. Block diagram of the proposed speech enhancement algorithm.

$\Psi[d(n)]$보다 값이 상당히 크기 때문에 $\Psi[d(n)]$는 생략될 수 있으며 최종적으로 식 (5)는 다음과 같이 나타낼 수 있다.

$$E\{\Psi[y(n)]\} \approx E\{\Psi[s(n)]\} \tag{6}$$

따라서 식(6)에 의해 잡음신호로 오염된 음성신호에 대한 TE의 기대 값은 깨끗한 음성신호에 대한 TE의 기대 값과 근사화 되므로 기존의 오염된 음성신호 $y(n)$에 기반한 피쳐 파라미터보다 TE operator에 의해 강화된 TE $\Psi[y(n)]$ 기반의 피쳐 파라미터들이 잡음신호에 대하여 보다 개선된 음성 특성을 도출할 수 있다.

Ⅲ. Teager Engergy와 음성부재확률을 적용하는 제안된 음성향상 알고리즘

Ⅱ장에서는 TE 기반의 피쳐 파라미터 도출을 위해 TE operator 대하여 간략히 설명하였다. Ⅲ장에서는 TE 기반의 SAP를 적용한 제안된 음성향상 알고리즘에 대하여 설명한다. 일반적으로 오염된 음성신호에 대한 통계적 모델 기반의 LR로 부터 주파수 밴드 별로 잡음과 음성의 특성이 확률 값으로 도출되는 LSAP는 잡음환경에서도 잡음에 대한 음성의 특성을 비교적 잘 나타내는 피쳐 파라미터로서 음성향상 알고리즘에서 잡음전력 갱신을 위한 VAD의 결정 식이나 스무딩(smoothing) 파라미터에 널리 적용되어 왔다. 하지만 LSAP 또한 잡음신호가 강한 낮은 SNR 환경에서는 잡음신호의 영향을 받아 잡음신호에 대한 음성신호의 특성이 약해져 정확한 VAD가 어려운 문제점이 발생한다.

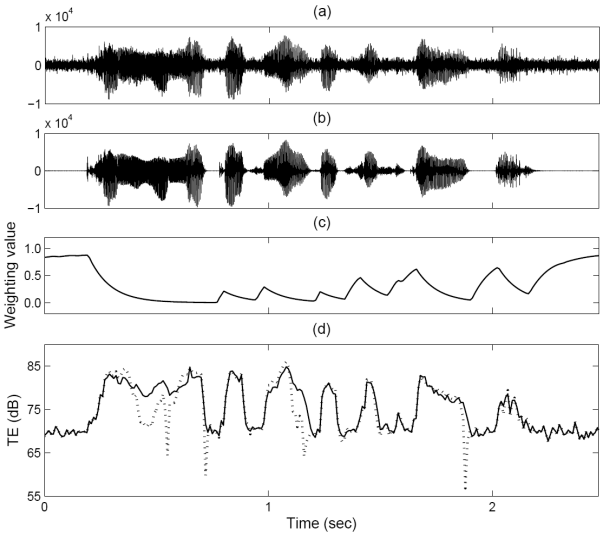

그림 2. 기존의 방법과 제안된 방법으로부터 도출된 TE (white noise, SNR=5 dB) (a) 오염된 음성신호 파형 (b) 깨끗한 음성신호 파형 (c) TEG를 적용한 Weighting parameter (d) 추정된 TE: 기존의 방법에 의한 TE (dotted line), 제안된 W_{TEG}에 의한 TE (solid line).
Fig. 2. TE derived by the conventional and proposed algorithm (white noise, SNR=5 dB) (a) Noisy speech waveform (b) Clean speech waveform (c) Weighting parameter based on TEG (d) Estimated TE: conventional method (dotted line), proposed W_{TEG} (solid line).

따라서 본 논문에서 다양한 잡음환경에서 효과적인 LSAP를 구하기 위해 다음과 같이 TE operator에 의해 잡음신호가 제거된 입력신호의 TE $\Psi[y(n)]$를 기반으로 LR을 도출하고 이로부터 TE 기반의 LSAP (TEL)를 음성향상을 위해 VAD의 피쳐 파라미터로 적용하는 방법 (VAD_{TEL})을 제안한다.

또한 기존의 TE operator에서는 상대적으로 SNR이 낮은 구간에서 TE operator에 의해 잡음신호가 제거되면서 음성신호 또한 상당히 제거되거나 왜곡되는 있는 문제가 발생할 수 있다. 따라서 본 논문에서는 이러한 문제점을 개선하기 위해 추정된 LR로부터 프레임 단위로 도출되는 TE 기반의 GSAP (TEG)를 개선된 TE를 구하기 위한 가중치 파라미터로 적용하는 알고리즘 (W_{TEG})을 제안한다. 그림 1은 제안된 음성향상 알고리즘의 블록도를 보여주고 있다. 첫째로, TE 기반의 LSAP를 추정하기 위해 오염된 음성신호의 TE $\Psi[y(n)]$에 대하여 음성신호가 존재하지 않을 때와 존재할 경우 각각의 가정 H_0, H_1을 다음과 같이 나타낼 수 있다 [6].

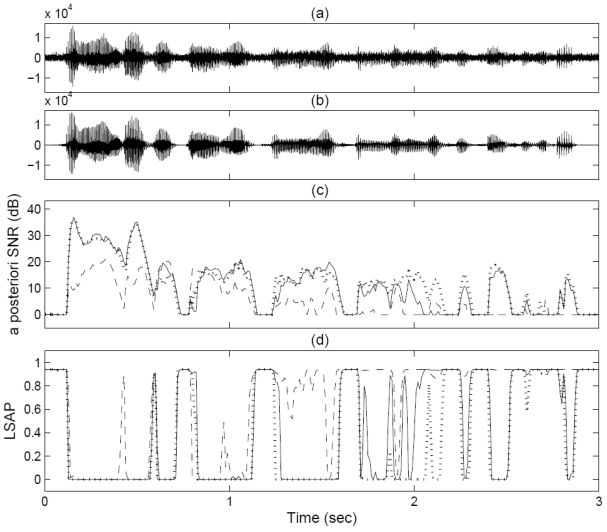

그림 3. 각각의 알고리즘에서 추정된 *a posteriori* SNR과 LSAP (white noise, SNR=5 dB) (a) 오염된 음성신호 파형 (b) 깨끗한 음성신호 파형 (c) 각각의 알고리즘에서 추정된 *a posteriori* SNR: 기존의 VAD_L (dashed line), VAD_{TEL} (solid line), $VAD_{TEL}+W_{TEG}$ (dotted line) (d) 각각의 알고리즘에서 도출된 LSAP: 기존의 VAD_L (dashed line), VAD_{TEL} (solid line), $VAD_{TEL}+W_{TEG}$ (dotted line).

Fig. 3. *a posteriori* SNR and LSAP estimated by the conventional and proposed algorithm (white noise, SNR=5 dB) (a) Noisy speech waveform (b) Clean speech waveform (c) Estimated *a posteriori* SNR: conventional VAD_L (dashed line), VAD_{TEL} (solid line), $VAD_{TEL}+W_{TEG}$ (dotted line) (d) Estimated LSAP: conventional VAD_L (dashed line), VAD_{TEL} (solid line), $VAD_{TEL}+W_{TEG}$ (dotted line).

$$H_0: \text{ speech absent } : \Psi[Y(i,k)] = \Psi[D(i,k)] \tag{7}$$

$$H_1: \text{ speech present } : \Psi[Y(i,k)] = \Psi[D(i,k)] + \Psi[S(i,k)]$$

여기서 $\Psi[Y(i,k)]$는 $\Psi[y(n)]$에 대해 이산 퓨리에 변환(DFT, discrete Fourier transform)을 통한 프레임 index i번째에서의 주파수 index k번째 주파수 성분을 나타내며 $\Psi[D(i,k)]$와 $\Psi[S(i,k)]$는 각각 주파수 영역에서 TE의 잡음과 음성신호에 대한 주파수 성분을 의미한다. 잡음과 음성신호가 통계적으로 독립이라고 가정하고 complex Gaussian 분포를 따른다는 가정하며 H_0와 H_1의 확률밀도함수는 다음과 같이 표현 된다 [6].

$$p(\Psi[Y(i,k)]|H_0) = \frac{1}{\pi\sigma_n(i,k)}\exp\left[-\frac{|\Psi[Y(i,k)]|^2}{\sigma_n(i,k)}\right] \tag{8}$$

$$p(\Psi[Y(i,k)]|H_1) = \frac{1}{\pi(\sigma_s(i,k)+\sigma_n(i,k))}\exp\left[-\frac{|\Psi[Y(i,k)]|^2}{\sigma_s(i,k)+\sigma_n(i,k)}\right] \tag{9}$$

여기서 $\sigma_s(i,k)$, $\sigma_n(i,k)$는 각각 TE 기반의 $\Psi[S(i,k)]$과 $\Psi[D(i,k)]$에 대한 전력을 나타내며 각 주파수 밴드 별 TEL를 구하기 위해 Bayes′ rule을 적용하면 다음과 같이 나타낼 수 있다 [6].

$$\begin{aligned} &p(H_0|\Psi[Y(i,k)]) \\ &= \frac{p(\Psi[Y(i,k)]|H_0)p(H_0)}{p(\Psi[Y(i,k)]|H_0)p(H_0)+p(\Psi[Y(i,k)]|H_1)p(H_1)} \\ &= \frac{1}{1+q_L\Lambda(\Psi[Y(i,k)])} \end{aligned} \tag{10}$$

여기서 $p(H_0)\,(=1-p(H_1))$은 음성 부재에 대한 *a priori* probability를 의미하고, $q_L(=p(H_1)/p(H_0))$는 0.0625로 설정되었으며 위의 식(8)와 식(9)을 식(10)에 대입하면 LR $\Lambda(\Psi[Y(i,k)])$는 다음과 같다.

$$\begin{aligned} \Lambda(\Psi[Y(i,k)]) &= \frac{p(\Psi[Y(i,k)]|H_1)}{p(\Psi[Y(i,k)]|H_0)} \\ &= \frac{1}{1+\zeta(i,k)}\exp\left[\frac{\eta(i,k)\zeta(i,k)}{1+\zeta(i,k)}\right] \end{aligned} \tag{11}$$

여기서, 파라미터로 $\eta(i,k)$, $\zeta(i,k)$는 각각 TE 기반의 *a posteriori* SNR과 *a priori* SNR로 아래와 같이 정의된다 [1].

$$\eta(i,k) \equiv \frac{|\Psi[Y(i,k)]|^2}{\sigma_n(i,k)} \tag{12}$$

$$\zeta(i,k) \equiv \frac{\sigma_s(i,k)}{\sigma_n(i,k)} \tag{13}$$

여기서 $\sigma_s(i,k)$는 음성신호에 대한 잡음전력을 의미하며 $\zeta(i,k)$을 추정하기 위해 Decision-Directed 추정 방법을 적용하였다[1]. 따라서 제안된 VAD 알고리즘은 식(10)로부터 도출된 TE 기반의 LSAP $p(H_0|\Psi[Y(i,k)])$에 적절한 문턱 값 T을 적용한 결정 식에 따른 음성검출 결과 f_{VAD}로 다음과 같이 나타낼 수 있다.

$$f_{VAD} = \begin{cases} \text{음성}(speech), & \text{if } p(H_0|\Psi[Y(i,k)]) < T \\ \text{비음성}(nonspeech), & \text{otherwise} \end{cases} \tag{14}$$

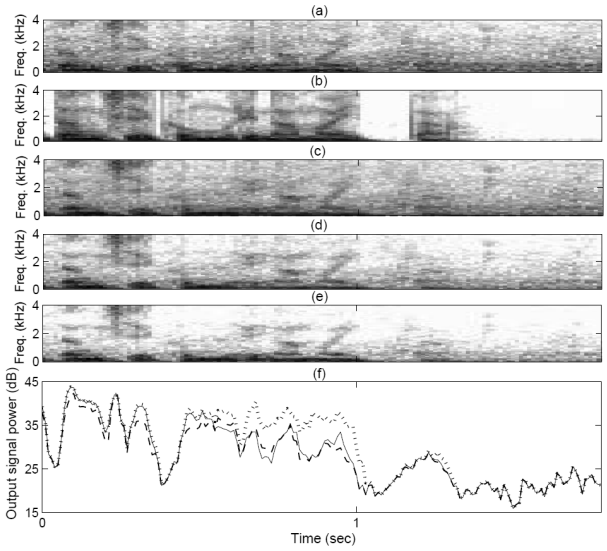

그림 4. 음성 스펙트로그램 (babble noise, SNR=5 dB) (a) 오염된 음성신호 (b) 깨끗한 음성신호 (c) 기존의 VAD_L 기반의 결과신호 (d) VAD_{TEL} 기반의 결과신호 (e) VAD_{TEL}+ W_{TEG} 기반의 결과신호 (f) 각각 알고리즘에 의해 도출된 결과신호의 전력: 기존의 방법 (dashed line), TE-LSAP (solid line), VAD_{TEL}+ W_{TEG} (dotted line).

Fig. 4. Speech spectrograms (babble noise, SNR=5 dB) (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the algorithm based on conventional VAD_L (d) Output signal obtained by the algorithm based on VAD_{TEL} (e) Output signal obtained by the algorithm based on VAD_{TEL}+W_{TEG} (f) Output signal power: conventional VAD_L (dashed line), VAD_{TEL} (solid line), VAD_{TEL}+W_{TEG} (dotted line).

여기서 T는 0과 1사이의 확률 값으로 표현되며 제안된 알고리즘에서는 테스트에 사용된 DB를 기반으로 $T=$ 0.9로 설정하였다.

또한, 제안된 음성향상 알고리즘에서는 TE operator의 성능 개선을 위하여 식 (2)를 TE-GSAP $p(H_0|\Psi[Y(i)])$ 기반의 가중치 파라미터를 프레임 index i를 적용하여 다음과 같이 나타낼 수 있다.

$$\Psi[\hat{Y}(i,k)] = F\{y(i,n)^2 - W_{GSAP}(i-1)y(i,n+1)y(i,n-1)\} \quad (15)$$

여기서 $\Psi[\hat{Y}(i,k)]$는 제안된 방법에 의해 추정된 TE의 주파수 성분을 의미하며 $F\{\bullet\}$는 주파수 변환을 위한 DFT 연산자를 나타낸다. 또한 $W_{GSAP}(i)$는 스무딩 파

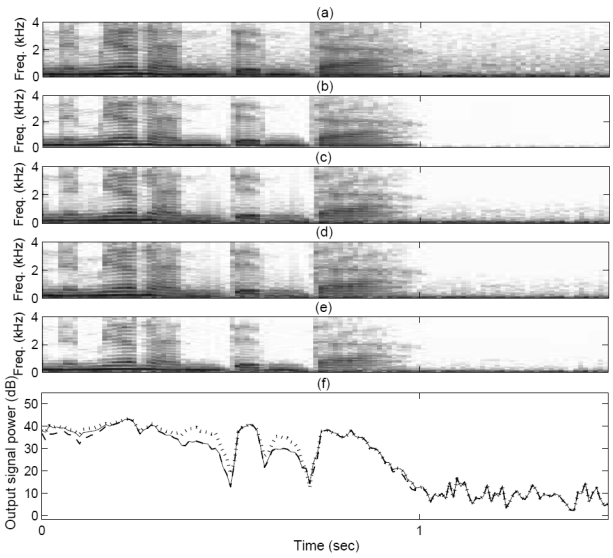

그림 5. 음성 스펙트로그램 (car noise, SNR=5 dB) (a) 오염된 음성신호 (b) 깨끗한 음성신호 (c) 기존의 VAD_L 기반의 결과신호 (d) TE-LSAP 기반의 결과신호 (e) VAD_{TEL}+ W_{TEG} 기반의 결과신호 (f) 각각 알고리즘에 의해 도출된 결과신호의 전력: 기존의 VAD_L 방법 (dashed line), VAD_{TEL} (solid line), VAD_{TEL}+ W_{TEG} (dotted line).

Fig. 5. Speech spectrograms (car noise, SNR=5 dB) (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the algorithm based on conventional VAD_L (d) Output signal obtained by the algorithm based on VAD_{TEL} (e) Output signal obtained by the algorithm based on VAD_{TEL}+ W_{TEG} (f) Output signal power: conventional VAD_L (dashed line), VAD_{TEL} (solid line), VAD_{TEL}+ W_{TEG} (dotted line).

라미터 $\alpha_g(=0.9)$에 의하여 스무딩 (smoothing) 된 TE-GSAP를 의미하며 다음과 같이 표현된다.

$$W_{GSAP}(i) = \alpha_g W_{GSAP}(i-1) + (1-\alpha_g)p(H_0|\Psi[Y(i)] \quad (16)$$

그리고 $p(H_0|\Psi[Y(i)])$는 식(11)에 의해 주파수 밴드 별로 도출된 LR로부터 다음과 같이 나타낼 수 있다[7].

$$\begin{aligned} & p(H_0|\Psi[Y(i)]) \\ &= \frac{p(H_0)\prod_{k=1}^{N} p(\Psi[Y(i,k)]|H_0)}{p(H_0)\prod_{k=1}^{N} p(\Psi[Y(i,k)]|H_0) + p(H_1)\prod_{k=1}^{N} p(\Psi[Y(i,k)]|H_1)} \\ &= \frac{1}{1+q_G\prod_{k=1}^{N}\Lambda(\Psi[Y(i,k)])} \end{aligned} \quad (17)$$

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	SNR (dB)	Method				
		VAD_L	VAD_G	VAD_{TEL}	VAD_{TEG}	$VAD_{TEL}+W_{TEG}$
White	5	1.999	2.130	2.026	2.139	2.045
	10	2.364	2.469	2.425	2.480	2.432
	15	2.702	2.784	2.770	2.788	2.770
Babble	5	2.350	2.340	2.359	2.348	2.367
	10	2.677	2.669	2.709	2.676	2.718
	15	2.959	2.971	3.005	2.973	3.009
Vehicle	5	3.304	3.444	3.342	3.454	3.346
	10	3.578	3.690	3.620	3.700	3.635
	15	3.842	3.913	3.883	3.919	3.891

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	SNR (dB)	Method				
		VAD_L	VAD_G	VAD_{TEL}	VAD_{TEG}	$VAD_{TEL}+W_{TEG}$
White	5	2.183	2.329	2.206	2.340	2.227
	10	2.606	2.719	2.668	2.731	2.678
	15	2.976	3.067	3.051	3.073	3.051
Babble	5	2.678	2.643	2.707	2.654	2.714
	10	3.047	3.027	3.087	3.031	3.098
	15	3.352	3.360	3.414	3.362	3.418
Vehicle	5	3.618	3.795	3.670	3.804	3.673
	10	3.935	4.073	3.992	4.079	4.006
	15	4.235	4.317	4.281	4.321	4.288

여기서 N은 각 프레임의 전체 주파수 index 개수를 나타내며 q_G는 적용한 테스트 DB를 기반으로 0.6으로 설정하였다.

그림 2 기존의 TE operator와 제안된 TE-GSAP 기반의 가중치를 적용한 방법에 의해 추정된 TE를 보여주고 있다. 그림 2의 (c)로부터 스무딩된 가중치 파라미터는 음성구간에서 0에 가까운 값을 보이고 비음성 구간에서는 1에 가까운 값을 나타내는 것을 볼 수 있다. 따라서 이러한 특성을 기반으로 제안된 방법의 식(15)에서는 음성구간에서 0에 가까운 가중치가 $y(i,n+1)y(i,n-1)$에 적용되고 에너지 $y(i,n)^2$에서 기존보다 감소된 값이 빼어짐으로서 결과적으로 기본의 방법에서보다 증가된 TE가 도출되고, 또한 비음성 구간에서 1에 가까운 가중치가 적용되어 기존에 가까운 TE 값이 도출된다. 따라서 제안된 방법에서는 음성 구간에서는 증간된 TE를 유도함으로써 기존의 방법보다 음성과 비음성의 특성을 강화시킬 수 있으며, 그림 2의 (d)로부터 깨끗한 음성신호 파형과 비교하여 볼 때 제안된 방법에 의한 TE가 기존의 방법보다 개선된 TE를 보여주고 있는 것을 알 수 있다.

최종적으로, 그림 3은 마이크로폰으로 입력된 오염된 음성신호에 대하여 기존의 LSAP를 VAD의 피쳐 파라미터로 적용한 알고리즘 (VAD_L)과 제안된 알고리즘에 대한 성능 비교를 위하여 추정된 a posteriori SNR과 LSAP를 보여주고 있다. 그림 3의 (c)로부터 오염된 음성신호 $y(n)$를 기반으로 도출된 기존의 VAD_L에서 추정된 a posteriori SNR보다 TE $\Psi[y(n)]$를 기반으로 한 TEL을 피쳐 파라미터로 사용하는 제안된 VAD_{TEL}

에 의해 추정된 a posteriori SNR이 잡음신호에 대하여 보다 뚜렷한 음성신호의 특성을 나타내는 것을 알 수 있으며, 또한 추가적으로 TE operator에 TEG를 가중치 파라미터로 적용한 W_{TEG} 방법을 추가적으로 적용한 $VAD_{TEL}+W_{TEG}$ 알고리즘이 상대적으로 SNR이 낮은 구간에서 발생할 수 있는 기존의 TE operator 방법에서 나타나는 단점을 개선하여 가장 효과적인 a posteriori SNR 추정치를 보여주는 것을 볼 수 있다. 또한 그림 3의 (d)로부터 SNR이 상대적으로 낮은 2-3초 구간에서 VAD_{TEL} 방법이 기존의 VAD_L 방법보다 향상된 LSAP를 보이는 것을 알 수 있으며 VAD_{TEL} 방법보다는 추가적으로 W_{TEG} 방법이 적용된 알고리즘에서 보다 개선된 LSAP를 추정치를 나타내는 것을 볼 수 있다.

최종적으로 식(14)의 제안된 VAD 알고리즘을 기반으로 음성향상 알고리즘에서 VAD 결과가 해당 구간을 비음성으로 결정하였을 경우 averaging rule을 통해 비음성 구간에서 잡음전력 $\lambda_d(i,k)$은 다음과 같이 업데이트 되어 추정된다.

$$\hat{\lambda}_d(i,k)=\alpha_d\lambda_d(i-1,k)+(1-\alpha_d)|Y(i,k)|^2 \quad (18)$$

여기서 α_d는 스무딩 파라미터를 의미하며 0.9로 설정하였다.

Ⅳ. 실험 및 결과고찰

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

$$C_{ovl}=1.594+0.805S_{PESQ}-0.512S_{LLR}-0.007S_{WSS} \quad (19)$$

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 8kHz로 샘플링 된 70개의 깨끗한 음성신호 수집하고 이를 세가지형태의 잡음을 다양한 SNR로 부가하여 오염된 음성신호를 생성하였으며 부가된 잡음은 NOISEX-92 데이터베이스의 white, babble, vehicle 잡음으로 SNR은 5, 10, 15 dB로 달리하였다. 또한 기존의 방법과 제안된 알고리즘의 성능평가는 minimum mean square error (MMSE) 기반의 잡음제거 이득을 가지는 음성향상 알고리즘에 적용하고 도출된 결과 음성신호에 대하여 평가하였다

그림 4와 그림 5는 각각의 알고리즘으로부터 도출된 결과신호에 대하여 음성 스펙트로그램과 결과신호의 전력 값을 보여주고 있다. 그림 4의 (d)와 (e) 그리고 그림 5의 (d)와 (e)로부터 TE 기반의 알고리즘이 적용된 음성향상 알고리즘으로부터 도출된 결과 음성신호가 기존의 방법에 보다 향상된 잡음제거 성능을 보이는 것을 볼 수 있다. 또한 그림 4의 (f)와 그림 5의 (f)로부터 제안된 알고리즘에 의한 결과신호가 음성구간에서 개선된 전력 값을 보이는 것을 알 수 있으며 특히 $VAD_{TEL}+W_{TEG}$에 의한 결과신호가 가장 향상된 전력 값을 나타내는 것을 볼 수 있다.

최종적으로 표 1과 표 2는 각각의 알고리즘에 의한 객관적인 음질평가를 보여주고 있으며 TE를 이용한 알고리즘에 대한 다양한 비교를 위하여 기존에 연구되었던 GSAP를 VAD의 피쳐 파라미터로 적용하는 방법 (VAD_G)과 TE 기반의 GSAP를 VAD를 위한 피쳐 파라미터로 적용한 알고리즘 (VAD_{TEG})의 성능 결과를 추가하였다 [10]. 표 1의 PESQ과 표2의 composite measure 수치로부터 모든 잡음환경에 대하여 비정상적(non-stationary)인 잡음 특성을 보이는 babble 잡음을 제외하고는 정상적인 (stationary) 잡음 특성을 보이는 대부분의 잡음환경에서 LSAP보다 GSAP를 VAD의 피쳐 파라미터로 적용한 알고리즘이 향상된 성능을 보이는 것을 알 수 있다. 이는 정상 잡음에서 결합된 형태의 LR로부터 도출된 GSAP가 통계적으로 LSAP 보다 많은 정보를 담고 있어 LSAP 보다 강인한 파라미터 특성을 보이는 것으로 추정 된다 [8]. 또한 기존의 LSAP와 GSAP 보다 TE를 적용한 방법이 개선된 결과 수치를 보였으며 최종적으로 제안된 $VAD_{TEL}+W_{TEG}$이 적용된 방법이 모든 잡음 환경에서 대하여 LSAP 기반의 VAD_L과 TEL 기반의 VAD_{TEL} 방법에 의해 도출된 평가 수치보다 향상된 결과를 보임으로서 향상된 성능을 나타내는 것을 볼 수 있다.

V. 결 론

본 논문에서는 잡음환경에서 효과적인 음성향상 알고리즘을 위해 TE 기반의 LR로부터 도출된 TEL을 잡음신호 갱신을 위한 VAD의 피쳐 파라미터로 이용하였으며 또한 프레임 별로 추정되는 TEG를 개선된 TE를 도출하기 위한 가중치 파라미터로 적용하였다. 제안된 알고리즘은 객관적인 테스트로부터 기존의 LSAP를 적용한 방법보다 향상된 결과를 나타내었다.

참 고 문 헌

[1] 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, Dec. 1984.

[2] 박윤식, 조규행, 장준혁, "복소 라플라시안 확률 밀도 함수에 기반한 음성 향상 기법," *전자공학회논문지*, 제44권, SP편 제6호, 111-117쪽, 2007. 11월.

[3] L. Karray, C. Mokbel and J. Monne, "Solutions for robust. speech/non-speech detection in wireless environment," *presented at the IVTTA*, Sep. 1988.

[4] 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.

[5] 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, Dec. 2008.

[6] 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.

[7] N. S. Kim and J.-H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, no. 5, pp. 108-110, May 2000.

[8] J. -H. Chang and N. S. Kim, "Speech enhancement : new approaches to soft decision," I*EICE Trans. Inf. and Syst.*, VolE84-D, No.9, pp 1231-1240, Sep. 2001.

[9] 조규행, 박윤식, 장준혁, "Smoothed global soft decision에 근거한 음성 향상 기법," *전자공학회논문지*, 제44권, SP편 제6호, 118-123쪽, 2007년 11월.

[10] 박윤식, 이상민, "잡음환경에서 Teager Energy 기반의 전역 음성부재확률을 이용하는 음성검출," *전자공학회논문지*, 제49권, SP편 제1호.

[11] Yi Hu and P. C. Loizou, "Evaluation of objective quality mea- sures for speech enhancement," *IEEE Trans. ASLP*, vol. 16, pp. 229 - 238, Jan. 2008.

저 자 소 개

박 윤 식(학생회원)

2006년 2월 인하대학교 전자공학과 학사
2008년 2월 인하대학교 전자공학부 석사
2008년 3월～현재 인하대학교 전자공학부 박사과정

<주관심분야 : 잡음제거, 음성검출, 음향학적 반향제거>

안 홍 섭(학생회원)

2010년 2월 인하대학교 전자공학과 학사
2012년 2월 인하대학교 전자공학과 석사
2012년 3월～현재 인하대학교 전자공학과 박사과정

<주관심분야 : 잡음제거, 보청기 알고리즘>

이 상 민(정회원)

1987년 인하대학교 전자공학과 학사 졸업
1989년 인하대학교 전자공학과 석사 졸업
2000년 인하대학교 전자공학과 박사 졸업

1989년 1월～1994년 7월 LG이노텍 선임연구원,
1995년 1월～2002년 3월 삼성종합기술원 책임 연구원,
2002년 4월～2005년 2월 한양대학교 의공학교실 연구교수,
2005년 3월～2006년 8월 전북대학교 생체정보공학부 조교수,
2006년 9월～현재 인하대학교 전자전기공학부 부교수

<주관심분야 : Healthcare system design, Psyco-acoustic, Brain-machine interface>