

논문 2012-49SP-1-12

# 잡음환경에서 Teager Energy 기반의 전역 음성부재확률을 이용하는 음성검출

(Voice Activity Detection Using Global Speech Absence Probability Based on Teager Energy in Noisy Environments)

박 윤 식\*, 이 상 민\*\*

(Yun-Sik Park and Sangmin Lee)

## 요 약

본 논문에서는 잡음환경에서 효과적인 음성을 검출하기 위한 새로운 음성 검출 (VAD, voice activity detection) 알고리즘을 제안한다. 통계적 모델에 기반의 Likelihood ratio (LR)를 통하여 도출되는 전역 음성부재확률 (GSAP, global speech absence probability)은 음성검출을 위한 피쳐 (feature) 파라미터로 널리 적용되고 있다. 하지만 신호 대 잡음 비 (SNR, signal-to-noise ratio)가 낮은 잡음환경에서는 정확한 GSAP 추정이 어려운 문제점을 가지고 있다. 따라서 제안된 방법에서는 잡음환경에서 강인한 VAD 알고리즘을 위하여 Teager energy (TE) 기반의 GSAP를 피쳐 파라미터로 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 voice activity detection (VAD) algorithm to effectively distinguish speech from nonspeech in various noisy environments. Global speech absence probability (GSAP) derived from likelihood ratio (LR) based on the statistical model is widely used as the feature parameter for VAD. However, the feature parameter based on conventional GSAP is not sufficient to distinguish speech from noise at low SNRs (signal-to-noise ratios). The presented VAD algorithm utilizes GSAP based on Teager energy (TE) as the feature parameter to provide the improved performance of decision for speech segments in noisy environment. 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**: 음향학적 반향 억제, 잡음 제거, 음성부재확률

## I. 서 론

음성검출기는 현재 모바일 음성통신 환경에서 다양한 어플리케이션과 결합되어 이슈가 되고 있는 음성인식 (speech recognition) 이나 잡음제거 (NS, noise suppression) 알고리즘과 같은 음성처리 시스템에 적용

되고 있으며 시스템 성능에 주요한 영향을 미치는 핵심 부분으로 인식되고 있다<sup>[1]</sup>. 특히, 직접적으로 음성검출 (VAD, voice activity detection) 알고리즘을 통해 잡음 신호를 추정하여 잡음을 제거하는 NS 알고리즘에서는 VAD 성능이 NS 알고리즘 성능에 전반적으로 영향을 주는 주요한 요소로 작용한다. 일반적으로 VAD 알고리즘은 음성 (speech)과 비음성 (nonspeech) 신호를 판별할 수 있는 피쳐 (feature) 파라미터를 도출하고 적절한 문턱 (threshold) 값을 피쳐 파라미터에 적용하는 결정식 (decision rule)의 형태로 음성과 비음성을 구분한다. 이러한 VAD에 사용되는 피쳐 파라미터로는 보편적으로 사용되는 스펙트럼 에너지 (spectral energy), ZCR

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

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

접수일자: 2011년7월5일, 수정완료일: 2011년9월14일

(zero-crossing ratio)에서부터 LPC (linear prediction coefficients) 및 통계적 모델에 기반한 likelihood ratio (LR) 등 다양한 피쳐 파라미터들이 적용되고 있다<sup>[2~3]</sup>. 하지만 VAD를 위해 적용되는 피쳐 파라미터들은 신호 대 잡음 비 (SNR, signal-to-noise ratio)가 큰 환경에서는 음성과 비음성에 대한 피쳐 파라미터 특성이 비교적 분명하지만 다양한 배경 잡음이 존재하는 실제 잡음 환경이나 SNR이 낮은 음성신호에 대해서는 피쳐 파라미터들이 잡음신호에 민감하기 때문에 VAD의 성능이 저하되는 문제점이 발생한다.

따라서 본 논문에서는 SNR이 낮은 잡음환경에서도 개선된 음성검출 성능을 위해 Teager energy (TE)<sup>[4]</sup>,<sup>[5]</sup>를 적용한 전역 음성부재확률 (GSAP, global speech absence probability) 기반의 VAD 알고리즘을 제안한다<sup>[6~8]</sup>. 구체적으로 제안된 방법에서는 잡음신호를 제거함으로써 잡음에 대한 음성의 특성을 강화시킬 수 있는 TE operator를 시간영역에서 적용한다. 또한 구해진 TE를 기반으로 주파수 영역에서 음성의 통계적 모델에 기반한 LR를 도출하고 이로부터 음성의 존재 유무가 확률 값으로 표현되는 GSAP를 VAD의 성능향상을 위한 피쳐 파라미터로 사용한다. 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 TE 기반의 GSAP를 도출하기 위한 TE operator에 대해 간략하게 설명한다. TE operator는 잡음신호를 제거하여 잡음신호에 대한 음성신호의 특성을 강화시켜줌으로써 잡음환경에서 효과적인 음성검출을 위해 적용되는 알고리즘으로서 잡음제거기나 잡음환경에서의 다양한 피쳐 파라미터들을 도출하는데 널리 사용되어 왔다<sup>[3~4]</sup>. 구체적으로 연속 (continuous) 시간에서의 신호를  $s(t)$ 라고 한다면 TE operator는 다음과 같이 정의된다<sup>[4~5]</sup>.

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

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

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

여기서  $n$ 은 이산 (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  $s(n)$ 과  $d(n)$ 은 상관관계가 없다는 가정하에 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 시간 영역에서의 TE를 의미하며  $s(n)$ 과  $d(n)$ 의 cross- $\Psi_d$  에너지  $\tilde{\Psi}_d[s(n), d(n)] = s(n)d(n) - 0.5s(n-1)d(n+1) - 0.5s(n+1)d(n-1)$ 이다. 그리고  $s(n)$ 과  $d(n)$ 은 zero mean이고 서로 독립이라 가정하면  $\tilde{\Psi}_d[s(n), d(n)]$ 의 기대 값은 0이 되고, 따라서 식 (4)는 다음과 같이 표현된다.

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

여기서 음성신호의 TE는 잡음신호의 TE보다 값이 상당히 크기 때문에 최종적으로 식 (5)은 다음과 같이 나타낼 수 있다<sup>[4~5]</sup>.

$$E\{\Psi_d[y(n)]\} \approx E\{\Psi_d[s(n)]\} \quad (6)$$

TE의 이러한 특성 때문에 기존의 보통 (regular) 에너지에 기반한 피쳐 파라미터 보다 TE 기반의 피쳐 파라미터들이 잡음환경에서 잡음신호에 대해 보다 개선된 음성 특성 도출할 수 있다.

## III. 제안된 Teager Energy 기반의 전역 음성부재확률을 이용하는 VAD 알고리즘

II장에서는 시간 영역에서 TE operator 대하여 간략히 설명하였다. III장에서는 TE 이용한 통계적 모델 기반의 전역 음성부재확률 (GSAP)를 피쳐 파라미터로 사용하는 새로운 VAD 알고리즘을 제안한다. 일반적으로

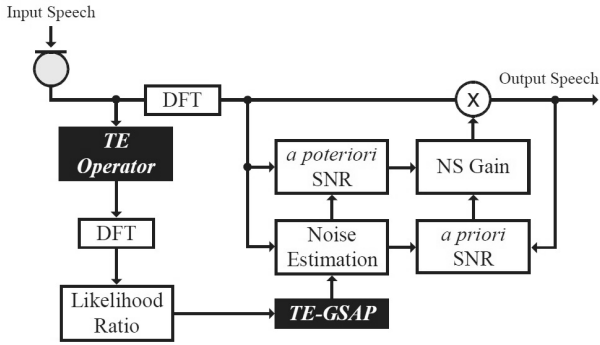


그림 1. 제안된 VAD 알고리즘의 블록도.  
Fig. 1. Block diagram of the proposed VAD algorithm.

통계적 모델에 기반의 LR에 의해서 프레임 단위로 잡음과 음성의 특성이 확률 값으로 도출되는 GSAP는 잡음환경에서도 음성과 비음성의 특성을 비교적 잘 나타내는 편이어서 다양한 VAD 알고리즘에 피쳐 파라미터로서 널리 사용되어 왔다<sup>[6-8]</sup>. 하지만 GSAP 또한 음성 신호에 비해 잡음신호가 강한 낮은 SNR 환경에서는 잡음신호의 영향을 받아 잡음신호에 대한 음성신호의 특성이 약해져 정확한 VAD가 어려운 문제점이 발생한다. 따라서 본 논문에서 잡음환경에서 효과적인 GSAP를 구하기 위해 시간 영역에서 TE operator에 의해 잡음신호가 제거된 입력신호에 대하여 TE 기반의 새로운 GSAP (TE-GSAP)를 도출한다. 그림 1은 제안된 VAD 알고리즘의 블록도를 보여주고 있으며 그림 1로부터 입력신호는 시간 영역에서 TE operator에 의해 TE를 도출하게 되고 DFT (discrete Fourier transform) 이후 주파수 영역에서 LR을 이용해 제안된 TE 기반의 TE-GSAP가 잡음추정에 적용되는 것을 볼 수 있다.

일반적으로 기존의 GSAP를 이용한 VAD 알고리즘은 GSAP에 적절한 문턱 값을 적용한 결정식 (decision rule)의 형태로 다음과 같이 표현된다.

$$f_{VAD} = \begin{cases} \text{음성 (speech)}, & \text{if } p(H_0|Y(i)) < T \\ \text{비음성 (nonspeech)}, & \text{otherwise} \end{cases} \quad (7)$$

여기서  $p(H_0|Y(i))$ 는 프레임 인덱스  $i$ 에서의 GSAP를 의미하며  $T$ 는 0과 1사이의 값을 가지는 문턱 값을 나타낸다<sup>[8]</sup>. 제안된 방법에서는 기존의 VAD 알고리즘에 제안된 TE 기반의 GSAP 적용하기 위하여 TE operator가 적용된 주파수 영역에서의 입력신호를 다음과 같이 도출한다.

$$Y_{TE}(i, k) = \Psi[Y(i, k)] \quad (8)$$

여기서  $\Psi[Y(i, k)]$ 는 TE 기반의 입력신호에 대한 시간 인덱스  $i$ 와 주파수 인덱스  $k$ 에서의 DFT 스펙트럼을 나타낸다.  $Y_{TE}(i, k)$ 로부터 제안된 TE-GSAP를 추정하기 위해 깨끗한 음성신호와 배경잡음이 상관관계가 없다는 가정 하에 음성신호가 존재하지 않을 때와 존재할 경우 각각의 가정  $H_0$ ,  $H_1$  을 다음과 같이 나타낼 수 있다.

$$H_0: \text{speech absent} : Y_{TE}(i, k) = D_{TE}(i, k) \quad (9)$$

$$H_1: \text{speech present} : Y_{TE}(i, k) = D_{TE}(i, k) + S_{TE}(i, k)$$

여기서 잡음신호  $D_{TE}(i, k)$ 는 음성신호인  $S_{TE}(i, k)$ 와 통계적으로 독립이고 음성과 잡음신호가 complex Gaussian 분포를 따른다는 가정하며  $H_0$ 와  $H_1$ 의 확률밀도함수는 다음과 같다<sup>[8]</sup>.

$$p(Y_{TE}(i, k)|H_0) = \frac{1}{\pi\lambda_n(i, k)} \exp\left[-\frac{|Y_{TE}(i, k)|^2}{\lambda_n(i, k)}\right] \quad (10)$$

$$p(Y_{TE}(i, k)|H_1) = \frac{1}{\pi(\lambda_s(i, k) + \lambda_n(i, k))} \exp\left[-\frac{|Y_{TE}(i, k)|^2}{\lambda_s(i, k) + \lambda_n(i, k)}\right] \quad (11)$$

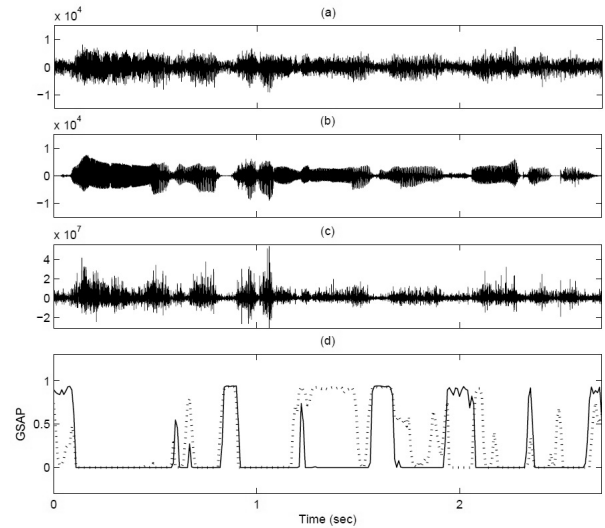


그림 2. GSAP 비교 (babble noise, SNR=0 dB) (a) 오염된 음성신호 파형 (b) 깨끗한 음성신호 파형 (c) Teager energy 파형 (d) GSAP: 기존의 방법 (점선), 제안된 방법 (실선).

Fig. 2. Comparison of GSAP (babble noise, SNR=0 dB) (a) Noisy speech waveform (b) Clean speech waveform (c) Teager energy waveform (d) GSAP: the conventional method (dashed line), the proposed method (solid line).

여기서  $\lambda_s(i, k)$ ,  $\lambda_x(i, k)$ 는 각각 음성과 잡음신호의 전력을 나타내며 각 주파수 채널별 SAP를 구하기 위해 Bayes' rule을 적용하면 다음과 같이 표현 된다<sup>[8]</sup>.

$$p(H_0|Y_{TE}(i, k)) = \frac{p(Y_{TE}(i, k)|H_0)p(H_0)}{p(Y_{TE}(i, k)|H_0)p(H_0) + p(Y_{TE}(i, k)|H_1)p(H_1)} \quad (12)$$

여기서  $p(H_0) (=1-p(H_1))$ 은 음성 부재에 대한 선행 확률 (*a priori probability*)이며, 각각의 주파수 채널별 성분이 통계적으로 독립이라는 가정으로부터, 식 (12)에서 제안된 TE-GSAP  $p(H_0|Y_{TE}(i))$ 는 아래와 같이 표현 된다<sup>[8]</sup>.

$$\begin{aligned} p(H_0|Y_{TE}(i)) &= \frac{p(H_0) \prod_{k=1}^N p(Y_{TE}(i, k)|H_0)}{p(H_0) \prod_{k=1}^N p(Y_{TE}(i, k)|H_0) + p(H_1) \prod_{k=1}^N p(Y_{TE}(i, k)|H_1)} \\ &= \frac{1}{1 + qA_{TE}(i)} \end{aligned} \quad (13)$$

여기서  $q = p(H_1)/p(H_0)$ 로 0.0625로 설정되었으며  $A_{TE}(i)$ 는 TE 기반의 전역 LR을 나타내며 아래와 같이 각 주파수 채널별 우도비의 곱으로 표현된다.

$$A_{TE}(i) = \prod_{k=1}^N A_{TE}(Y_{TE}(i, k)) = \prod_{k=1}^N \frac{p(Y_{TE}(i, k)|H_1)}{p(Y_{TE}(i, k)|H_0)} \quad (14)$$

여기서  $N$ 은 프레임 별 전체 주파수 밴드 개수를 의미하며 채널별 우도비  $A_{TE}(Y_{TE}(i, k))$ 는 식 (11)과 식 (12)로부터 다음과 같이 유도된다<sup>[8]</sup>.

$$\begin{aligned} A_{TE}(Y_{TE}(i, k)) &= \frac{p(Y_{TE}(i, k)|H_1)}{p(Y_{TE}(i, k)|H_0)} \\ &= \frac{1}{1 + \xi(i, k)} \exp \left[ \frac{\gamma(i, k)\xi(i, k)}{1 + \xi(i, k)} \right] \end{aligned} \quad (15)$$

여기서, 파라미터로  $\gamma(i, k)$ ,  $\xi(i, k)$ 는 각각 *a posteriori* SNR과 *a priori* SNR로 아래와 같이 정의 된다.

$$\gamma(i, k) \equiv \frac{|Y_{TE}(i, k)|^2}{\lambda_n(i, k)} \quad (16)$$

$$\xi(i, k) \equiv \frac{\lambda_s(i, k)}{\lambda_n(i, k)} \quad (17)$$

최종적으로 식 (13)의 TE 기반 GSAP  $p(H_0|Y_{TE}(i))$ 에 적절한 문턱 값  $\Gamma_{TE}(=0.8)$ 을 적용하여 제안된 VAD 알

고리즘은 다음과 같이 표현된다.

$$f_{VAD} = \begin{cases} \text{음성 (speech)}, & \text{if } p(H_0|Y_{TE}(i)) < \Gamma_{TE} \\ \text{비음성 (nonspeech)}, & \text{otherwise} \end{cases} \quad (18)$$

그림 2의 (c)는 마이크로 폰으로 오염된 입력신호에 대하여 TE operator를 적용한 파형을 보여주고 있다. 그림 2의 (c)부터 TE는 오염된 입력신호보다 상대적으로 잡음은 상당한 제거되고 음성신호는 잡음신호에 비해 크기가 증가한 것을 볼 수 있다. 또한 오염된 음성신호에 대하여 기존의 GSAP와 제안된 TE-GSAP의 비교 그림을 보여주고 있는 그림2의 (d)로부터 제안된 방법에 의한 TE-GSAP가 잡음환경에서 기존의 GSAP보다 개선된 피쳐 파라미터 특성을 나타내는 것을 알 수 있다. 잡음제거기에서는 이러한 VAD 알고리즘을 기반으로 VAD 결과가 해당 구간을 비음성으로 결정하였을 경우 averaging rule을 통해 비음성 구간에서 잡음 전력  $\lambda_d(i, k)$ 를 다음과 같이 업데이트 한다.

$$\lambda_d(i, k) = \alpha_d \lambda_d(i-1, k) + (1 - \alpha_d) |Y(i, k)|^2 \quad (19)$$

여기서  $\alpha_d$ 는 스무딩 (smoothing) 파라미터를 의미하며 0.9로 설정하였다.

#### 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)<sup>[9]</sup> 및 다음과 같이 전반적인 음질  $C_{ovl}$ 로 표현되는 composite measure 테스트를 실시하였다<sup>[10]</sup>.

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

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 8kHz로 샘플링 된 60개의 깨끗한 음성신호 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 제안된 TE-GSAP 기반 VAD의 성능평가는 제안된 알고

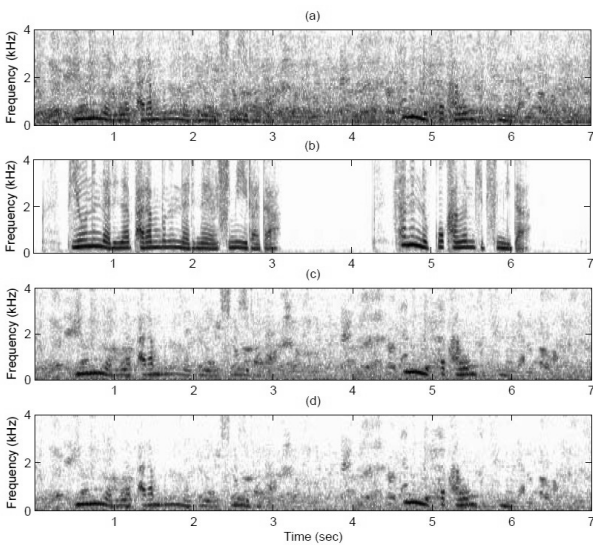


그림 3. 음성 스펙트로그램 (babble noise, SNR=0 dB)  
 (a) 오염된 음성신호 (b) 깨끗한 음성신호  
 (c) 기존의 GSAP 기반의 결과신호 (d) 제안된 TE-GSAP 기반의 결과신호

Fig. 3. Speech spectrograms (babble noise, SNR=0 dB)  
 (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the method based on conventional GSAP (d) Output signal obtained by the method based on proposed TE-GSAP.

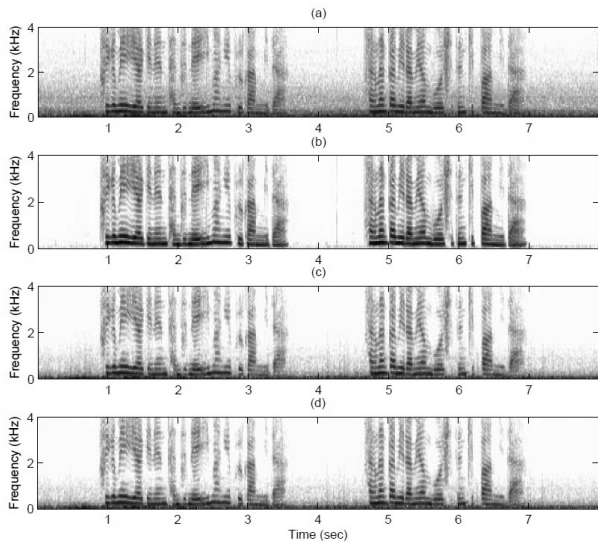


그림 4. 음성 스펙트로그램 (vehicle noise, SNR=0 dB)  
 (a) 오염된 음성신호 (b) 깨끗한 음성신호 (c) 기존의 GSAP 기반의 결과신호 (d) 제안된 TE-GSAP 기반의 결과신호

Fig. 4. Speech spectrograms (vehicle noise, SNR=0 dB)  
 (a) Input noisy signal (b) Clean speech (c) Output signal obtained by the method based on conventional GSAP (d) Output signal obtained by the method based on proposed TE-GSAP.

표 1. 다양한 잡음환경에서 기존과 제안된 TE-GSAP 알고리즘에 대한 PESQ 수치 .

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

Noise type	SNR (dB)	Method	
		GSAP	TE-GSAP
White	0	1.650	1.657
	5	2.074	2.085
	10	2.423	2.434
	15	2.755	2.759
Babble	0	1.886	1.901
	5	2.273	2.280
	10	2.615	2.618
	15	2.927	2.929
Vehicle	0	3.122	3.134
	5	3.416	3.428
	10	3.685	3.696
	15	3.910	3.917

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

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

Noise type	SNR (dB)	Method	
		GSAP	TE-GSAP
White	0	1.794	1.802
	5	2.299	2.311
	10	2.707	2.720
	15	3.077	3.083
Babble	0	2.128	2.149
	5	2.591	2.603
	10	2.993	2.997
	15	3.346	3.348
Vehicle	0	3.467	3.480
	5	3.803	3.813
	10	4.102	4.109
	15	4.349	4.354

리즘을 minimum mean square error (MMSE)<sup>[11]</sup> 기반의 잡음제거 이득을 가지는 잡음제거기에 적용하고 도출된 결과 음성신호에 대하여 평가하였다.

그림 3과 그림 4는 잡음환경에서 각각의 VAD 알고리즘이 잡음제거기에 적용되어 도출된 결과 음성신호에 대한 음성 스펙트로그램을 보여주고 있다. 그림 3의 (d)와 그림 4의 (d)로부터 제안된 TE-GSAP 기반의 VAD 알고리즘이 적용된 잡음제거기가 기존의 알고리즘보다 향상된 잡음제거 성능을 보이는 것을 볼 수 있다. 또한 각각의 알고리즘에 의한 PESQ 수치를 보여주고 있는

표 1로부터 모든 잡음환경에 대하여 제안된 알고리즘이 기존의 알고리즘보다 향상된 PESQ 수치를 보여 주고 있는 것을 알 수 있다. 최종적으로 표 2는 composite measure에 대해 수치를 보여주고 있으며 표 2로부터 제안된 알고리즘이 기존의 방법보다 모든 잡음 환경에서 개선된 composite measure 결과를 보여 주고 있는 것을 볼 수 있다. 따라서 잡음신호가 음성신호에 비해 비교적 작은 SNR=15 dB의 잡음환경에서부터 잡음신호의 의해 상당히 오염된 환경인 SNR=0 dB까지의 다양한 잡음환경에서 모든 잡음 종류에 대해 표 1과 표 2에서 객관적 음질평가에 의해 제안된 알고리즘이 기존의 방법보다 향상된 음질평가 수치를 보임으로서 제안된 VAD 알고리즘이 기존의 방법보다 개선된 성능을 보인다는 것을 알 수 있다.

## V. 결 론

본 논문에서는 잡음환경에서 효과적인 VAD 알고리즘을 위해 TE를 적용한 GSAP 기반의 새로운 VAD 방법을 제안하였다. TE operator에 의해서 잡음이 제거된 입력신호를 기반으로 TE 기반의 GSAP를 도출하여 기존의 GSAP보다 잡음환경에서 개선된 피쳐 파라미터 특성을 보였으며 객관적인 테스트로부터 제안된 알고리즘이 기존의 방법보다 향상된 결과를 나타내었다.

## 참 고 문 헌

- [1] L. Karray, C. Mokbel and J. Monne, "Solutions for robust. speech/non-speech detection in wireless environment," presented at the IVTTA, Sep. 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, May 1977.
- [3] 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, Jan. 1999.
- [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] 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, Apr. 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] 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.
- [9] 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, May 2001.
- [10] Yi Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. ASLP*, vol. 16, pp. 229 - 238, Jan. 2008.
- [11] 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.

— 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>



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,  
 Psycho-acoustic, Brain-machine interface>