

비정상 잡음환경에서 음질향상을 위한 적응 임계치 알고리즘

Adaptive Threshold for Speech Enhancement in Nonstationary Noisy Environments

이 수 정*, 김 순 협**
(Soo-Jeong Lee*, Sun-Hyob Kim**)

*광운대학교 음성신호처리연구실, **광운대학교 컴퓨터공학과
(접수일자: 2008년 7월 31일; 수정일자: 2008년 9월 8일; 채택일자: 2008년 10월 8일)

본 논문에서는 비정상 잡음환경에서 음질향상을 위한 새로운 방법을 제안한다. 정상 잡음환경에서 음질향상을 위한 잡음제거 방법으로 주파수 차감법이 잘 알려져 있다. 그러나 실제 잡음환경은 대부분 비정상적인 특성을 나타낸다. 제안한 방법은 다양한 잡음과 비정상 환경에서 잘 동작 할 수 있도록 적응 임계치를 위한 자동제어 파라미터를 사용한다. 특히, 자동제어 파라미터는 *a posteriori* SNR를 이용한 선형함수를 적용하여 잡음레벨의 증감에 따라 적응 임계치를 제어한다. 제안한 알고리즘은 음질향상을 위해 Hangover (HO)을 이용한 주파수 차감법과 결합한다. 알고리즘의 성능은 다양한 잡음환경에서 ITU-T P.835 signal distortion (SIG)와 segment signal to-noise ratio (SNR)로 평가하여 (HO)을 이용한 음성검출과 minimum statistics (MS) 방법에 비해 우수한 결과를 나타냈다.

핵심용어: 잡음예측, 음질향상, 적응 임계치, 비정상 잡음환경

투고분야: 음성처리 분야 (2,3)

This paper proposes a new approach for speech enhancement in highly nonstationary noisy environments. The spectral subtraction (SS) is a well known technique for speech enhancement in stationary noisy environments. However, in real world, noise is mostly nonstationary. The proposed method uses an auto control parameter for an adaptive threshold to work well in highly nonstationary noisy environments. Especially, the auto control parameter is affected by a linear function associated with an *a posteriori* signal to noise ratio (SNR) according to the increase or the decrease of the noise level. The proposed algorithm is combined with spectral subtraction (SS) using a hangover scheme (HO) for speech enhancement. The performances of the proposed method are evaluated ITU-T P.835 signal distortion (SIG) and the segment signal to-noise ratio (SNR) in various and highly nonstationary noisy environments and is superior to that of conventional spectral subtraction (SS) using a hangover (HO) and SS using a minimum statistics (MS) methods.

Keywords: Noise estimator, Speech enhancement, Adaptive threshold, Non-stationary noisy environment

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

I. 서론

잡음예측 및 제거방법은 다양한 음성통신 기기분야의 중요한 핵심기술이다. 잡음에 오염된 음성신호의 음질과 명료도 향상을 통하여 음성통신시스템의 성능을 향상시킬 수 있다. 현재 가장 잘 알려진 잡음제거 기술 중

주파수 차감법 (SS) [1] 있다. 이 방법은 간단하면서도 음성왜곡 및 잔여잡음을 제어할 수 있어 많은 분야에 응용되고 있다. 주파수 차감법에서는 음성의 시작과 끝부분을 묵음으로 추정하는 음성검출방법 voice active detector (VAD) [2]을 사용하지만, 실제 잡음환경과 낮은 (SNR)에서는 시스템의 성능이 급격히 감소된다. 최근 기존의 음성검출방법 (VAD)이 아닌 비정상 잡음환경에 적용할 수 있는 새로운 알고리즘들이 활발히 연구·발표되고 있다 [3-8]. 대표적인 방법으로 minimum statistics (MS) [3]

책임저자: 이 수 정 (leesoo86@kw.ac.kr)
서울시 노원구 월계동 광운대학교 음성신호처리연구실
(전화: 02-940-5123; 팩스: 02-941-5189)

이 있지만 MS 방법 역시 잡음레벨이 급격히 변하는 비정상 잡음환경에서 잡음의 최소레벨 검색을 위한 응답시간 지연 문제로 알고리즘의 성능이 저하되는 단점을 가지고 있다 [4][7-8].

본 논문에서는 비정상 잡음환경에서 효과적인 새로운 적응예측 알고리즘을 제안한다. 제안한 방법은 시간 및 주파수영역의 noisy 파워 스펙트럼의 분산과 그 값의 정규화 비율 기반 위에 [8] 다양한 잡음과 비정상 환경에서 잘 동작 할 수 있도록 자동제어 파라미터를 이용한다. 특히, 자동제어 파라미터는 *a posteriori* SNR을 적용한 선형함수를 사용하여 *a posteriori* SNR 증가할 경우 제어파라미터 값이 감소되고, *a posteriori* SNR 감소할 경우 제어파라미터 값이 증가되어 비정상 잡음환경과 같은 잡음레벨의 급격한 변화에 잘 적응할 수 있도록 하였다. 제안된 잡음예측 알고리즘을 기존의 HO 방법을 사용한 주파수 차감 법과 결합시켜 음성외곡은 최소화하고 음질의 성능은 향상시켰다. 새로운 알고리즘의 성능은 다양한 잡음 환경에서 ITU-T P.835 signal distortion (SIG) [9]와 segment (SNR) 그리고 비정상 환경에서 개선된 음성신호 스펙트럼에 의해 평가하여 HO+SS 및 MS+SS에 비해 향상된 결과를 나타냈다 [2][3].

II. MS와 VAD 기반 잡음예측

이 장에서는 잡음파워 예측을 위한 대표적인 방법으로 비정상 잡음환경에서 사용하는 MS 방법과 기존에 잘 알려진 VAD 방법을 간단히 설명한다.

2.1. MS 기반 잡음파워 예측

본 논문에서 음성신호와 잡음신호를 무상관으로 가정한다. 잡음에 오염된 음성신호 $y(n)$ 는 $y(n) = x(n) + d(n)$ 로 표현되며, $x(n)$ 은 음성신호, $d(n)$ 은 잡음신호를 나타낸다. 잡음에 오염된 음성신호 $y(n)$ 을 Hamming window 함수를 이용하여 중첩된 구간들로 나누고, 단 구간 Fourier 변환 (STFT) [1]을 사용하여 시간 및 주파수 축으로 나타낸다.

$$|Y(l,k)|^2 \approx |X(l,k)|^2 + |D(l,k)|^2 \quad (1)$$

여기서 $(l=1,2,\dots,L)$ 는 프레임 (즉, 시간 축 인덱스)이며, $(k=1,2,\dots,K)$ 주파수축 인덱스이다. 수식 (1)은 잡음

에 오염된 음성파워스펙트럼을 나타내며 $|X(l,k)|^2$ 는 음성파워스펙트럼 그리고 $|D(l,k)|^2$ 는 잡음 파워스펙트럼이다 [1].

MS 알고리즘의 잡음예측방법은 음성의 묵음구간 최소치를 사용한다. 이 값은 음성부재 구간에서 잡음수준까지 감소하므로 식 (2)을 이용하여 잡음의 파워를 예측할 수 있다 [3].

$$P(l,k) = \alpha P(l-1,k) + (1-\alpha) |Y(l,k)|^2 \quad (2)$$

여기서 $P(l,k)$ 는 지수 평활 법을 이용해 구한 평활된 잡음파워스펙트럼이며, $\alpha(0.7 \leq \alpha < 0.85)$ 는 평활 상수이다. 평활상수 α 값이 1에 가까운 값을 가질 경우 스펙트럼의 평활정도가 증가되며, 음성 peaks가 확장되어 작은 notches는 제거되므로 부정확한 잡음예측을 유도할 수 있다 [3][8]. 이러한 단점을 해결하기 위해 최적화된 가변 평활 파라미터는 아래 수식 (3)로 나타낸다 [3].

$$\alpha_{opt}(l,k) = \frac{1}{1 + (P(l-1,k)/\bar{\sigma}_d^2(l,k) - 1)^2} = \frac{1}{1 + (\gamma - 1)^2} \quad (3)$$

여기서 $\bar{\gamma} = P(l-1,k)/\bar{\sigma}_d^2(l,k)$ 는 평활된 *a posteriori* SNR을 나타내며, $\alpha_{opt}(l,k)$ 는 최대치 $\alpha_{max} = 0.96$ 로 제한하여 $\bar{\gamma} = P(l-1,k)/\bar{\sigma}_d^2(l,k)$ 가 1일 때 발생하는 교착상태의 문제를 해결하고, 음성구간에 포함된 잡음을 고려하여 최소치는 $\alpha_{min} = 0.3$ 으로 제한한다. 위의 수식(3)에서 구한 가변 평활파라미터에 의해 구한 파워스펙트럼의 최소치는 잡음파워의 평균보다 작은 값을 갖는 바이어스 상태이므로 바이어스보상을 이용하여 최적화된 잡음 파워의 예측치를 구한다 [8].

$$\hat{\sigma}_d^2(l,k) = B_{min}(l,k) \cdot P_{min} \quad (4)$$

여기서 $B_{min}(l,k)$ 는 바이어스 보상함수, $P_{min}(l,k)$ 는 최소잡음파워이다. 바이어스 보상함수는 잡음에 오염된 문장을 기반으로 하며 아래와 같이 근사식으로 나타낸다 [3].

$$B_{min}(l,k) \approx 1 + (C-1) \frac{2}{Q_q(l,k)} \quad (5)$$

식(5)에서 C는 프레임 수이며, Q_{eq} 는 수식 (6)로 표현 된다.

$$Q_{eq} \approx \frac{2\sigma_d^4(l-1, k)}{\widehat{var}(P(l, k))} \quad (6)$$

여기서 $\widehat{var}(P(l, k))$ 는 평활 된 파워 스펙트럼 $P(l, k)$ 의 분산이며, 수식(4)의 $P_{min}(l, k)$ 는 수식 (7)을 이용하여 구한다 [3].

$$P_{min}(l, k) = \min\{P(l, k), P(l-1, k), \dots, P(l-C, k)\} \quad (7)$$

2.2. MS 방법의 문제점 분석

서론에서 언급한 MS 알고리즘의 잡음추정 방법은 한 정된 구간의 특정 윈도우에 대한 오염된 음성신호의 파워 스펙트럼의 최소치를 이용하여 얻을 수 있다 [3]. 그림 1과 같은 비정상 잡음환경에서 실험결과 잡음수준이 급격히 증가하는 경우 (즉, SNR 감소) 최소치 검색에 따른 응답시간 지연으로 인하여 잡음수준 예측치가 바이어스 되는 문제점을 가지며, 또한 바이어스 보상합수를 사용한 후에도 전체적인 잡음수준이 실제잡음수준에 비해 작은 값을 갖는 바이어스 상태를 나타내고 있다. 그림 1에서 car noise SNR 15 (dB) 다음에 car noise SNR 5 (dB)로 잡음레벨이 증가하는 구간에서 응답시간지연으로 바이어스 상태가 생기는 현상을 볼 수 있다.

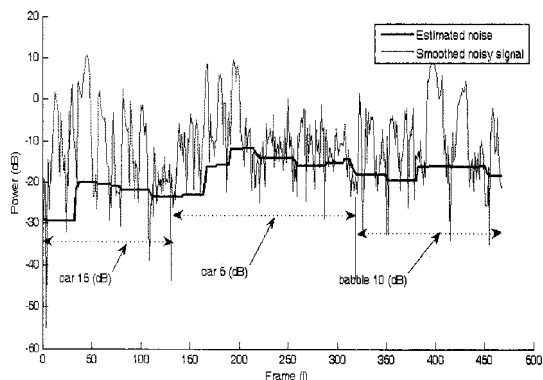


그림 1. 비정상 잡음환경 car noise 15 (dB) and 5 (dB), babble noise 10 (dB)에서 MS 알고리즘의 오염된 음성 파워스펙트럼과 잡음추정

Fig. 1. Noisy power spectrum and noise estimate of MS method for car noise 15 (dB) and 5 (dB), babble noise 10 (dB) in a nonstationary at f=625 Hz.

2.3. 스펙트럼 거리와 HO를 이용한 음성검출

이 절에서는 음성신호를 검출하는 간단한 방법으로 스펙트럼의 차이를 이용한 음성검출방법(VAD)을 살펴 본다 [2].

$$S_{dis} = 20 \cdot (\log_{10}(Y(l, k)) - \log_{10}(\widehat{D}(k))) \quad (8)$$

위의 수식 S_{dis} 는 각 주파수 인덱스의 스펙트럼의 차이를 나타낸다. 아래 수식 (9)은 음수 값을 제거 하기위해 사용한다.

$$S_{dis} = \text{MAX}(S_{dis}, 0) \quad (9)$$

$$Dis = \frac{1}{K} \sum_{k=1}^K S_{dis} \quad (10)$$

그림 2의 n_m 는 임계치 = 3을 실험치로 사용하며, $hangover_L$ ($6 \leq L \leq 8$)는 행 오버 길이를 나타낸다. 행 오버의 길이는 프레임 길이, 오버랩 길이 그리고 샘플링 주파수에 따라 결정 된다 [2].

$$\widehat{D}(k) = \frac{\widehat{D}(k) \cdot noise_L + |Y(l, k)|^2}{noise_L + 1} \quad \text{if } speech = 0 \quad (11)$$

그림 2에서 음성 flag가 reset 되면 식(11)에 의해 추정된 잡음신호가 갱신된다. 여기서 $noise_L$ 는 $noise_L = hangover_L + 1$ 로 나타낸다. 그러나 스펙트럼거리를 이용한 음성검출방법은 잡음수준이 안정적인 구간에서는 잘 동작하지만, 잡음수준이 급격히 변하는 비정상 잡음 환경 및 SNR (dB)이 낮은 구간에서는 알고리즘의 성능이 급격히 저하된다.

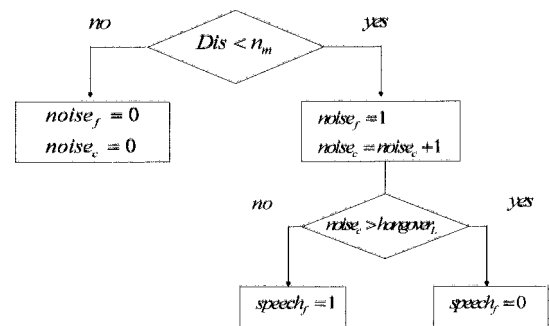


그림 2. 음성과 비 음성 결정을 위한 행 오버 블록도

Fig. 2. Hangover block diagram for decision of speech and non-speech.

III. 제안한 잡음예측 알고리즘

제안한 잡음예측알고리즘은 주파수 도메인에서 잡음에 오염된 음성 파워 스펙트럼의 분산을 이용하여 비정상 잡음환경에서도 잡음 임계치를 잘 추적할 수 있도록 적응적 기법을 사용하며, 음성과 잡음의 전이 구간에서 생기는 음성왜곡을 최소화 하기위해 기존의 행오버 방법과 결합시켜 기존방법 MS에 비해 간단하지만 성능에서 우수한 알고리즘을 제안한다. 제안한 알고리즘의 성능평가를 위해 주파수 차간 변 [1]과 결합시켜 잡음제거 후 음질향상정도를 평가한다.

$$\mu_l = \frac{1}{L} \sum_{l=1}^L |Y(l, k)|^2, \quad (12)$$

여기서 μ_l 는 각 주파수 구간의 잡음에 오염된 음성과 파워 스펙트럼의 평균을 각 프레임으로 나타낸다.

$$\sigma_l = \frac{1}{L} \sum_{l=1}^L (|Y(l, k)|^2 - \mu_l)^2 \quad (13)$$

$$\hat{\sigma} = \frac{1}{L} \sum_{l=1}^L \sigma_l \quad (14)$$

수식 (13)은 프레임 축의 오염된 잡음 파워 스펙트럼의 분산이며, 수식 (14)은 수식 (13)을 이용하여 얻은 정규화 값으로 시간-주파수 축의 잡음파워의 추정 값으로 사용한다.

$$\eta = \frac{\sigma_l}{\hat{\sigma}} \quad (15)$$

수식 (15)는 시간과 주파수 축에서 noisy 파워 스펙트럼의 분산과 정규화 값의 비율로 결정된다 [7]. 결과적으로 식 (15)의 값 (즉, 분산과 정규화 비율)이 적응 임계치보다 큰 값을 가질 경우 음성구간으로, 작은 값을 가질 경우 비 음성구간으로 추정한다. 일반적으로 음성이 포함된 구간은 발생된 음성스펙트럼의 영향으로 음성이 포함되지 않은 구간에 비해 매우 큰 분산을 가진다. 상대적으로 음성이 포함되지 않은 구간은 매우 작은 분산을 (0과 가까운 값) 나타낸다. 이런 특성을 이용하여 음성구간과 비 음성구간을 구분한다. 아래 그림 3은 수식 (15)의 비율과 적응 임계치를 보여준다.

3.1. A posteriori (SNR)을 이용한 선형함수

제안한 자동제어 파라미터는 a posteriori SNR을 이용한 선형함수를 적용하여 a posteriori SNR 증가할 경우 제어 파라미터 값이 감소하고, a posteriori SNR 감소할 경우 제어 파라미터 값이 증가하여 잡음레벨의 급격히 변화는 비정상 잡음환경에서도 잘 적용할 수 있도록 하였다.

$$SNR_l = 10 \cdot \log \left(\frac{\text{norm}(|Y(l, k)|^2, 2)}{\text{norm}(\hat{D}(k), 2)} \right), \quad (16)$$

여기서 $\hat{D}(k) = \frac{1}{5} \sum_{l=1}^5 |Y(l, k)|^2$ 는 초기 복음구간의 프레임의 평균치이며, norm은 $\text{norm} = \sqrt{(D_1^2 + D_2^2 + \dots + D_K^2)}$ 나타낸다.

$$\delta_s = (\delta_{\min} - \delta_{\max}) / (SNR_{\max} - SNR_{\min}), \quad (17)$$

$$\delta_{\text{off}} - \delta_{\max} = \delta_s \cdot SNR_{\min}. \quad (18)$$

여기서 δ_s 는 δ 를 구하기 위한 기울기 함수이며, 식(18)은 offset이다. 식(17)과 (18)에 사용된 상수 $\delta_{\min} = 0.1$, $\delta_{\max} = 0.3$ 는 다양한 잡음환경에서 반복된 실험을 통해 얻어진 실험값을 사용하고, $SNR_{\max} = 20\text{dB}$ 을 $SNR_{\min} = -5\text{dB}$ 을 사용한다 [10].

$$\delta = \delta_s \cdot SNR_l + \delta_{\text{off}}, \quad (19)$$

여기서 식(17), (18), 그리고 (19)를 이용하여 제어 파라미터 δ 를 구한다. 결과적으로 a posteriori SNR과 선형함수를 이용하여 제어 파라미터 δ 가 조절된다. 그림 4

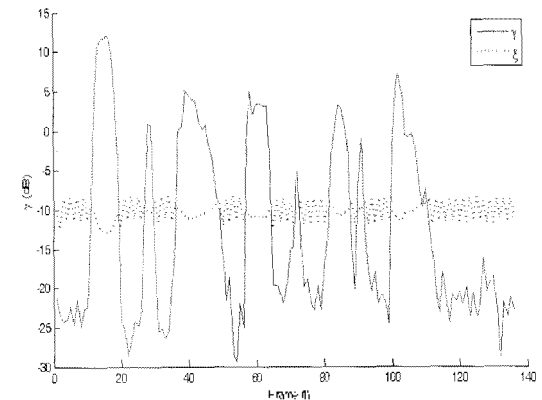


그림 3. 자동차 잡음 환경 SNR 5 (dB)에서 ratio (실선)와 적응 임계치(점선)

Fig. 3. Ratio (solid line) and adaptive threshold (dot line) in car noise 5 (dB) SNR environment.

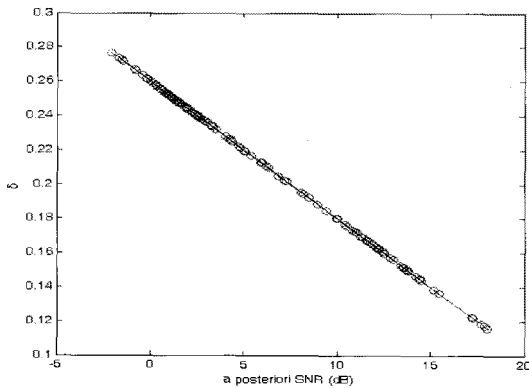


그림 4. 제어변수 δ 를 위한 *a posteriori* SNR
Fig. 4. The *a posteriori* SNR for the control parameter δ .

에서 *a posteriori* SNR이 증가하면, 제어 파라미터 δ 가 감소하는 것을 보여준다. 잡음 레벨이 $SNR_{max} = 20\text{ dB}$ 에서 $SNR_{min} = -5\text{ dB}$ 사이의 값을 가진 경우 제어파라미터 δ 값을 변화시켜 임계치를 조절한다.

3.2. 비정상 잡음환경을 위한 적응임계치 알고리즘

적응임계치 알고리즘은 3.1 절에서 구한 제어 파라미터 δ 에 따라 적응임계치를 제어한다. *A posteriori* SNR이 증가하면, 제어변수 δ 가 감소하며, 그 결과로 적응임계치 수준이 감소된다. 그와 반대로, *a posteriori* SNR이 감소하면, 제어변수 δ 가 증가하고, 수준도 증가한다. 그림 5에서는 비정상 잡음환경에서 잡음수준을 추적하는 임계치를 볼 수 있다. 임계치가 식(15)에서 구한 ratio에 따라 변화하는 것을 볼 수 있다.

$$\xi_1 = \gamma_1 + \delta \quad (20)$$

여기서 ξ_1 은 적응 임계치 초기 값으로 식(15)의 γ_1 과 앞 절에서 구한 제어 파라미터 δ 을 이용하여 구한다.

$$\xi_U = \xi_1 \cdot 2.0 \quad (21)$$

$$\xi_L = \xi_1 \cdot 0.5 \quad (22)$$

식(21)의 ξ_U 는 적응 임계치 상한 값으로 적응 임계치의 초기 값 식(20)에 SNR (dB) 기준 3 dB 증가 값인 상수 2를 곱하여 상한 값으로 사용하며, 임계치 상한 값보다 큰 값의 ratio는 음성구간으로 간주된다. 식(22)는 ξ_L 는 적응 임계치 하한 값으로 적응 임계치의 초기 값 수식(20)에 SNR (dB) -3 dB 감소 값인 상수 0.5를 사용한다.

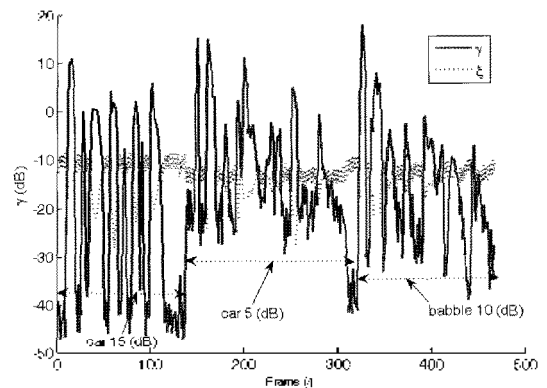


그림 5. 비정상 잡음환경 car noise (15 dB와 5 dB), babble noise (10 dB)에서 제안한 알고리즘의 오염된 음성 파워스펙트럼의 적응 임계치 추정

Fig. 5. Adaptive thresholds estimation on the time index for car noise (15 dB and 5 dB), babble noise 10 (dB) in a nonstationary noisy environments.

$$\alpha_t = \xi_{t-1} - \gamma_{t-1} \quad (23)$$

식(23)은 이전 구간의 적응임계치와 ratio를 이용하여 현재 적응 임계치를 추적하는 변수로 사용한다.

$$\begin{aligned} &\text{if } \alpha_t > \delta, \\ &\quad \xi_t = \xi_{t-1} \cdot (1 - C_a) + \xi_L \cdot C_a \\ &\text{else if } 0.0 \leq \alpha_t < \delta, \\ &\quad \xi_t = \xi_{t-1} \\ &\text{else} \\ &\quad \xi_t = \xi_{t-1} \cdot (1 - C_a) + \xi_U \cdot C_a \end{aligned} \quad (24)$$

적응 임계치 초기 값은 묵음구간(즉, 비 음성)으로 잡음만 존재한다고 가정한다. 만약 특정구간이 임계치의 하한 값보다 작은 값을 가진 경우 해당구간은 비 음성구간으로 가정하고, 임계치의 상한 값보다 큰 값을 가진 경우 음성구간으로 가정한다. 추적변수 α_t 의 값이 제어 파라미터 δ 보다 큰 값을 가질 경우, 하한 임계치에 가중치 상수 C_a ($0.5 \leq C_a < 1$)을 곱하고, 이전구간의 적응 임계치에 $(1 - C_a)$ 을 곱하여 이전구간 적응 임계치에 비해 감소된 값을 현재구간의 적응 임계치 값으로 사용한다. 결론적으로 적응 임계치는 상한과 하한 임계치 사이에서 정규화 비율을 추적한다. 위에서 사용한 가중치 상수는 반복 실험을 통하여 얻어진 값이다. 또한 가중치 상수를 통하여 적응임계치의 기울기를 조절할 수 있다.

표 2. 제안한 음성 향상 알고리즘과 (PROPOSED)과 기존 방법 (HOSS)와 (MSSS) 방법의 SIG 수치 비교

Table 2. The SIG result for the proposed speech enhancement method and conventional methods.

noise	method	SIG		
		5	10	15
white	HO+SS	2.58	2.76	2.37
	MS+SS	1.65	3.13	3.26
	Proposed	2.86	3.40	4.06
babble	HO+SS	2.95	3.43	3.48
	MS+SS	3.22	3.14	3.67
	Proposed	3.19	3.75	3.81
car	HO+SS	3.45	3.56	3.59
	MS+SS	2.69	3.75	3.89
	Proposed	3.66	3.85	3.91
nonstationary	HO+SS		3.81	
	MS+SS		3.86	
	Proposed		3.88	

3.3. 적응 임계치 알고리즘과 HO 결합

잡음예측알고리즘을 행 오버 방법과 결합하여 정상 및 비정상 잡음환경에 적용한다. 행 오버 방법은 음성신호의 파워가 약한 부분 (음성전이 부분)이 잡음구간으로 처리되어 생기는 음성왜곡을 보상한다 [2].

그림 6은 그림 2에서 D_{fs} 와 η_m 을 η 와 ξ 로 대체하여 사용한다.

$$\eta_s = (\eta_{min} - \eta_{max}) / (SNR_{max} - SNR_{min}), \quad (25)$$

$$\eta_{off} = \eta_{max} - \eta_s \cdot SNR_{min}, \quad (26)$$

여기서 η_s 는 η 를 구하기 위한 기울기 함수이며, 식(26)은 offset이다. 식(25)과 (26)에서 사용된 상수 $\eta_{min} = 1$, $\eta_{max} = 3$, $SNR_{max} = 20\text{ dB}$ 그리고 $SNR_{min} = -5\text{ dB}$ 을

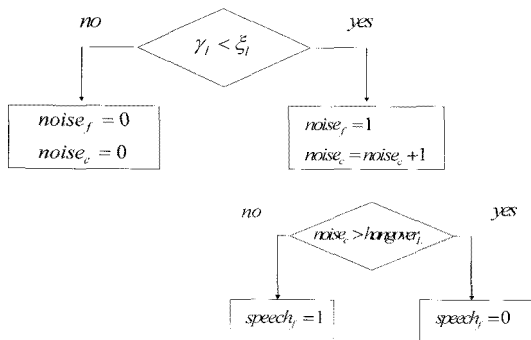


그림 6. 음성과 비 음성 결정을 위한 개선된 행 오버 블록도
Fig. 6. Updated Hangover block diagram for decision of speech and non-speech.

사용한다 [2].

$$\eta_k = \eta_s \cdot SNR_k + \eta_{off}, \quad (27)$$

식(27) 오버차감파라미터를 나타내며, $SNR_k = 10 \cdot \log(\hat{Y}(k,l)^2 / \hat{D}(k))$ 는 주파수 구간에서 *a posteriori* SNR을 나타낸다. η_k 는 주파수 차감 시 음성신호의 왜곡에 영향을 미친다 [1][10]. 식(28)의 $\beta(0 < \beta \ll 1)$ 는 spectral floor로 잔여잡음을 제어한다 [1]. 마지막으로, 개선된 음성 파워 스펙트럼을 $\hat{X}(k,l)^2$ 주파수 차감 법 [1]을 이용하여 식(28)과 같이 나타낸다.

$$|\hat{X}(k,l)|^2 = \max[Y(k,l)^2 - \eta_k \cdot \hat{D}_k \cdot \beta \cdot \hat{D}_k] \quad (28)$$

개선된 음성신호는 식(28)을 이용하여 inverse STFT를 이용하여 구한다.

V. 실험 결과 및 고찰

본 논문에서 제안된 알고리즘의 성능평가를 위해 white, babble, 과 car noise 환경에서 객관적 테스트를 수행하였다. 실험에 사용된 DB는 남성과 여성 각 2명이 발생한 30개의 음성샘플을 NOIZEUS 110에서 사용하였고, 음성신호를 8 kHz로 샘플링 하고 (SIFT)를 사용하여 50% overlapping Hamming windows 256 samples를 이용하였다. 표 1은 기존방법 및 제안한 방법의 segmental SNR과

표 1. 다양한 잡음환경에서 segmental SNR (dB)
Table 1. The average result of segmental SNR (dB).

noise	method	SNR (dB)		
		5	10	15
white	HO+SS	9.4	11.83	12.92
	MS+SS	7.76	11.68	12.86
	Proposed	9.76	12.13	13.72
babble	HO+SS	7.96	11.11	12.95
	MS+SS	5.18	10.13	14.56
	Proposed	8.5	11.77	15.54
car	HO+SS	8.07	11.77	12.98
	MS+SS	5.53	9.91	14.69
	Proposed	9.16	12.2	15.99
nonstationary	HO+SS		0.54	
	MS+SS		0.42	
	Proposed		1.05	

(SIG) 을 보여주고 있다 [5-6]. 표 1의 결과로부터 제안한 방법이 기존의 방법보다 잡음제거에서 향상된 결과를 나타냈다.

또한, 제안한 알고리즘의 음성향상 성능을 평가하기 위해 남성, 여성의 상이한 문장을 다양한 잡음과 SNR (dB) 에서 실험하였고, 잡음제거 후 음질향상의 성능을 평가하였다. 또한 극심한 비정상 잡음환경의 실험을 위해 남성과 여성 그리고 각기 다른 특성과 SNR (dB)을 갖는 음성신호를 결합하여 실험환경을 설정하였다. 그림 7의 (c) 와 (d)에서 잡음 SNR이 car 15 (dB)에서 SNR car 5 (dB)로 변하는 2.2초 이후 잔여잡음을 볼 수 있고, (d)의 제안한 방법은 기존방법을 이용한 방법에 비해 비정상 잡음 환경에서도 잘 적응하여 잡음을 제거한 것을 볼 수 있다.

표 2는 잡음환경에서 음질왜곡 결과를 보여준다. 음질 왜곡정도에 따라 (왜곡이 전혀 없는 경우=5, 거의 없는 경우=4, 약간의 왜곡=3, 심각한 왜곡=2, 그리고 매우 심각한 왜곡=1)으로 평가되었다. 제안한 방법이 babble 5 (dB) SNR을 제외하고 기존에 방법에 비해 우수한 결과를 나타내었다.

VI. 결론

본 논문에서는 시간-주파수 도메인 기반의 분산의 정규화 기법 기반 위에 다양한 잡음과 비정상 환경에서 작동 할 수 있도록 자동제어 파라미터를 사용하였다. 이 파라미터는 *a posteriori* SNR을 이용한 선형함수를 적용하여 *a posteriori* SNR 증가할 경우 제어 파라미터 값이 감소되고, *a posteriori* SNR 감소할 경우 제어 파라미터 값이 증가되어 비정상 잡음환경과 같은 잡음레벨의 급격한 변화에도 잘 적응할 수 있도록 하였다. 제안된 잡음예측 알고리즘을 기존의 HO 방법을 사용한 주파수차감법과 결합시켜 음성향상의 성능을 평가하였다. 새로운 알고리즘의 성능은 다양한 잡음환경에서 ITU-T P.835 (SIG) 와 segment (SNR), 그리고 개선된 음성신호에 의해 평가하여 기존의 방법에 비해 향상된 결과를 나타냈다. 특히 비정상 유색잡음 환경에서는 향상된 성능을 보여주었다.

감사의 글

본 연구는 광운대학교 2008년도 교내연구비 지원에 의해 수행되었습니다.

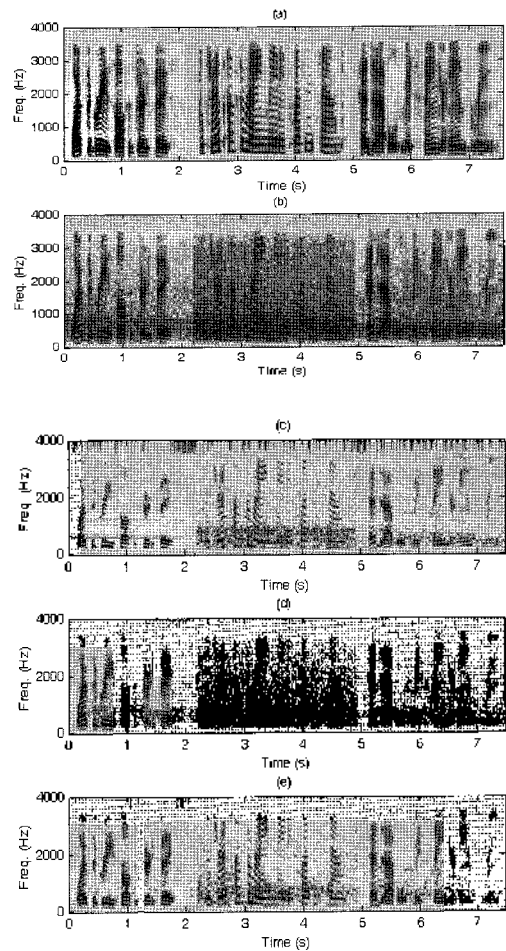


그림 7. Car noise 15 dB, car noise 5 dB, babble noise 10 dB의 비정상 잡음환경에서 음성강화 알고리즘의 결과. (a) 음성신호, (b) 잡음에 오염된 음성신호, (c) HOSS를 사용한 음성강화, (d) MSSS 를 사용한 음성강화 (e) 제안한 방법을 이용한 음성강화.

Fig. 7. Frequency domain results of speech enhancement for car noise at 15 dB, car noise at 5 dB, babble noise at 10 dB SNR in a nonstationary environments. (a) Original speech; (b) Noisy speech; (c) Enhancement speech using SSHO; (d) Enhancement speech using SSMS; (e) Enhancement speech using Proposed.

참고 문헌

1. M. Berouli, M. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", Proc. IEEE Int. Conf. Acoust. Speech Signal Processing, 208-211, 1979.
2. A. Davis and S. Nordholm, "A Low Complexity Statistical Voice Activity Detector with Performance Comparisons to ITU-T/ ETSI Voice Activity Detectors, Proc. IEEE Int. Conf. Information Communi. & Signal Process., 119-123, 2003.
3. R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", IEEE Trans. Speech Audio Processing, 9(5), 504-512, 2001.
4. I. Cohen, "Noise Spectrum in Adverse Environments: Improved Minima Controlled Recursive Averaging", IEEE Trans. Speech and Audio Proc., 11(5), 466-475, 2003.
5. L. LIN, W.H. HOLMES, and E. AMBIKAIKAJAH, "Adaptive noise

- estimation algorithm for speech enhancement", *ELECTRONICS LETTERS*, 39(9), 754-755, 2003.
6. R. SUNDARRAJAN, C.L. PHILIPOS, "A noise-estimation algorithm for highly non-stationary environment", *SPEECH COMMUNICATION*, 48, 220-231, 2006.
 7. S. J. Lee and S. H. Kim, "Noise Suppression Using Normalized Time-Frequency Bin Average and Modified Gain Function for Speech Enhancement in Nonstationary Noisy Environments", *THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA*, 27(1E), 1-10, 2008.
 8. 박윤식, 정준혁, "강인한 음성향상을 위한 Minimum Statistics 와 Soft Decision의 확률적 결합의 새로운 잡음전력", *한국음향학회지*, 26(4), 153-158, 2007.
 9. ITU-T, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm", *ITU-T Recommendation P.835*, 2003.
 10. C. L. PHILIPOS, *SPEECH ENHANCEMENT (Theory and Practice)*, 1st edition, (CRC Press, Boca Raton, Fl., 2007), pp.495-498.

저자 약력

•이 수 정 (Soo-Jeong Lee)

한국음향학회지 Vol.26 No.3F 참조

•김 순 협 (Sun-Hyob Kim)

한국음향학회지 Vol.26 No.3E 참조