

음질향상을 위해 비선형 함수와 사전 음성부재확률을 이용한 최소통계법의 잡음전력편의 보상방법

Noise-Biased Compensation of Minimum Statistics Method using a Nonlinear Function and *A Priori* Speech Absence Probability for Speech Enhancement

이 수 정*, 이 강 성**, 김 순 협***

(Soo-Jeong Lee*, Gang-Seong Lee**, and Sun-Hyob Kim***)

*성균관대학교 정보통신공학부 BK21 사업단, **광운대학교 교양학부, ***광운대학교 컴퓨터공학과
(접수일자: 2008년 6월 30일; 수정일자: 2008년 9월 4일; 채택일자: 2008년 11월 26일)

본 논문에서는 비정상 잡음환경에서 음질향상을 위한 비선형 함수와 사전 음성부재 확률을 이용한 최소 통계치 (MS) 방법의 잡음전력편의 보상 방법을 제안한다. 비정상 잡음환경에서 잡음전력추정을 위해 최소 통계치 방법이 잘 알려져 있지만, 예측된 잡음전력 추정 값은 실제 잡음 전력 값보다 하향 편이 되는 특성을 나타낸다. 제안한 방법은 비선형 함수를 적용한 적응보상파라미터와 사전 음성부재 확률 값을 혼용하는 잡음전력편의 보상방법이다. 특히, 적응보상 파라미터는 사후 SNR 을 이용한 비선형함수를 적용하여 잡음수준의 증감에 따라 파라미터 값을 조절한다. 또한, 사전 음성부재확률 값이 1로 수렴할 경우, 적응보상파라미터 값은 각 수파수별로 최대치까지 증가하지만, 확률 값이 0에 가까워지면 반대의 특성을 나타낸다. 제안한 알고리즘의 잡음전력추정 및 음질향상의 성능평가를 위해 다양한 종류의 잡음과 비정상적인 극심한 잡음환경을 설정하여 실험하고, 음질향상을 위해 수파수 차감법과 결합하였다. 알고리즘의 성능은 다양한 잡음환경의 신호 대 잡음비 (SNR) 와 Itakura-Saito 유사왜곡 평가 법 을 이용하여 기존 최소 통계치 (MS) 방법에 비해 우수한 결과를 나타냈다.

핵심용어: 잡음예측, 음질향상, 적응 임계치, 비정상 잡음환경

투고분야: 음성처리 분야 (2,3)

This paper proposes a new noise-biased compensation of minimum statistics(MS) method using a nonlinear function and *a priori* speech absence probability(SAP) for speech enhancement in non-stationary noisy environments. The minimum statistics(MS) method is well known technique for noise power estimation in non-stationary noisy environments. It tends to bias the noise estimate below that of true noise level. The proposed method is combined with an adaptive parameter based on a sigmoid function and *a priori* speech absence probability (SAP) for biased compensation. Specifically, we apply the adaptive parameter according to the *a posteriori* SNR. In addition, when the *a priori* SAP equals unity, the adaptive biased compensation factor separately increases to δ_{max} in each frequency bin, and vice versa. We evaluate the estimation of noise power capability in highly non-stationary and various noise environments, the improvement in the segmental signal-to-noise ratio (SNR), and the Itakura-Saito Distortion Measure (ISDM) integrated into a spectral subtraction (SS). The results shows that our proposed method is superior to the conventional MS approach.

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

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

I. 서론

잡음추정 및 음질향상방법은 다양한 음성인식, 음성통신 기기분야 와 인공지능시스템의 중요한 요소기술이다. 잡음에 오염된 음성신호의 음질과 명료도 향상을 통하여 음성통신응용시스템들의 성능을 향상시킬 수 있다. 현재 가장 잘 알려진 잡음추정 방법은 음성의 시작과 끝부분을 묵음으로 추정하는 음성검출방법 voice active detector (VAD) [1]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 [2-9]. 대표적인 방법으로 최소 통계치 방법 (MS) [2]이 있지만, MS 방법 역시 잡음레벨이 변하는 실제 잡음환경에서 잡음전력추정 값이 실제 잡음 전력 값보다 하향 편향 되는 특성을 나타낸다. 이런 결과로 잔여잡음이 생기는 문제점을 가지고 있고 [3][6-7], 잡음전력편향의 문제를 해결하기 위한 방법이 발표되고 있다 [8].

본 논문에서도 이러한 단점을 해결하기 위해 비정상 잡음환경에서도 효과적인 변형된 최소 통계치 알고리즘을 제안한다. 제안한 방법은 비선형 함수를 적용한 적응보상파라미터 와 사전 음성부재 확률 값을 혼용하는 잡음전력편의 보상방법이다. 특히, 적응보상 파라미터는 사후 SNR 을 이용한 비선형함수를 적용하여 잡음수준의 증감에 따라 파라미터 값을 조절한다. 또한, 사전 음성부재확률 값이 1로 수렴할 경우 적응보상파라미터 값은 각 주파수별로 최대치까지 증가하며, 확률 값이 0과 가까워지면 반대의 특성을 나타낸다. 제안한 알고리즘의 잡음전력추정을 위해 다양한 종류의 잡음과 비정상적인 극심한 잡음환경을 설정하여 실험하고, 음질향상을 위해 주파수 차감법과 결합하였다. 알고리즘의 성능은 다양한 잡음환경에서 신호 대 잡음비 (SNR)와 Itakura-Saito 음질왜곡 평가 방법을 이용하여 기존 minimum statistics (MS) 방법에 비해 우수한 결과를 나타냈다.

II. 최소 통계치 (MS) 기반 잡음예측

이 장에서는 잡음파워 예측을 위한 대표적인 방법으로 비정상 잡음환경에서 사용하는 MS [3] 방법을 간단히 설명한다.

2.1. MS 기반 잡음파워 추정

본 논문에서는 음성신호와 잡음신호를 무상관으로 가정하

다. 잡음에 오염된 음성신호는 $y(n) = x(n) + d(n)$ 로 표현되며, $x(n)$ 은 음성신호, $d(n)$ 은 잡음신호를 나타낸다. 잡음에 오염된 음성신호 $y(n)$ 는 Hamming window 함수를 이용하여 중첩된 구간들로 나누고, 단 구간 Fourier 변환 (STFT)을 사용하여 시간 및 주파수 축으로 나타낸다.

$$|Y(l,k)|^2 = |X(l,k)|^2 + |\hat{D}(l,k)|^2 \quad (1)$$

여기서 $(l = 1, 2, \dots, L)$ 는 프레임(즉, 시간 축 인덱스)이며, $(k = 1, 2, \dots, K)$ 주파수축 인덱스 이다. 수식 (1)은 잡음에 오염된 음성파워스펙트럼을 나타내며 $|X(l,k)|^2$ 는 음성파워스펙트럼 그리고 $|\hat{D}(l,k)|^2$ 는 잡음 파워스펙트럼이다 [8].

MS 알고리즘의 잡음추정방법은 음성의 묵음구간의 최소치를 사용한다. 이 값은 음성부재 구간에서 잡음수준까지 감소하므로 식 (2)를 이용하여 잡음의 파워를 예측할 수 있다 [2][7].

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

여기서 $P(l,k)$ 는 지수 평활 법을 이용해 구한 평활된 잡음파워스펙트럼이며, $\alpha (0.7 \leq \alpha \leq 0.85)$ 는 평활 상수이다. 평활상수 α 값이 1에 가까운 값을 가진 경우 스펙트럼의 평활정도가 증가되며, 음성 peaks가 확장되어 작은 notches 는 제거되므로 부정확한 잡음예측을 유도할 수 있다 [2][7]. 이러한 단점을 해결하기 위해 최적화된

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

가변 평활 파라미터는 아래 수식 (3)으로 나타낸다.

여기서 $\bar{\gamma} \approx P(l-1,k)/\sigma_d^2(l,k)$ 는 평활된 a posteriori SNR 을 나타내며, $\alpha_{opt}(l,k)$ 는 최대치 $\alpha_{max} = 0.96$ 로 제한하여 $\bar{\gamma} \approx P(l-1,k)/\sigma_d^2(l,k)$ 가 1일 때 발생하는 교착상태의 문제를 해결하고, 음성구간에 포함된 잡음을 고려하여 최소치는 $\alpha_{min} = 0.3$ 으로 제한한다. 위의 수식 (3)의 가변 평활파라미터를 이용해서 얻은 파워스펙트럼의 최소치는 잡음파워의 평균보다 작은 값을 갖는 바이어스 상태이므로 바이어스보상을 이용하여 최적화된 잡음 파워의 예측치를 구한다 [7][11].

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

여기서 $B_{min}(l,k)$ 는 바이어스 보상함수, $P_{min}(l,k)$ 는

최소잡음과워이다. 바이어스 보상함수는 잡음에 오염된 분산을 기반으로 하며 아래와 같이 근사식으로 나타낸다 [2].

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

식 (5)에서 C=40은 프레임 수이며, Q_{eq} 는 수식 (6)으로 표현된다.

$$Q_{eq} = \frac{2\hat{\sigma}_d^2(l-1, k)}{\widehat{\text{var}}(P(l, k))} \quad (6)$$

여기서 $\widehat{\text{var}}(P(l, k))$ 는 평활된 파워 스펙트럼 $P(l, k)$ 의 분산이며, 수식 (4)의 $P_{\min}(l, k)$ 는 수식 (7)을 이용하여 구한다 [2].

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

2.2. MS 방법의 문제점 분석

서론에서 언급한 MS 알고리즘의 잡음추정 방법은 한정된 구간의 특정 윈도우에 대한 오염된 음성신호의 파워 스펙트럼의 최소치를 이용하여 얻을 수 있다 [2]. 그림 1과 같은 극심한 비정상 잡음환경에서는 잡음수준이 급격히 증가하는 경우 (즉, SNR 감소) 최소치 검색에 따른 응답시간 지연으로 인한 잡음수준 추정치가 하향 바이어스 되는 문제를 나타내며, 또한 바이어스 보상함수를 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.

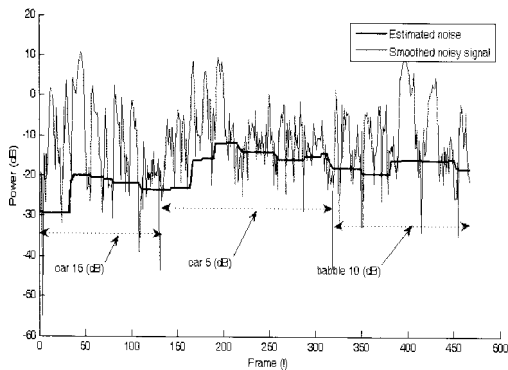


그림 1. 비정상 잡음환경 car noise 15 (dB) and 5 (dB), babble noise 10 (dB)에서 MS 알고리즘의 오염된 음성 파워스펙트럼과 잡음추정

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.

III. 사전 음성부재 확률 (SAP) 추정

이 절에서는 음성의 존재와 부재를 두 가지 상태모델로 아래와 같이 나타낸다 [10].

$$\begin{aligned} H_0(k, l) : Y(k, l) &= D(k, l) \\ H_1(k, l) : Y(k, l) &= X(k, l) + D(k, l) \end{aligned} \quad (8)$$

위의 수식 $H_0(k, l)$ 와 $H_1(k, l)$ 는 음성부재와 음성존재를 나타내며, 잡음신호 $D(k, l)$ 는 음성신호 $X(k, l)$ 와 통계적 독립으로 가정한다. 음성과 잡음신호가 복소 가우시안 분포라는 가정으로 음성부재확률을 아래와 같이 표현한다 [9].

$$P(H_0(k, l)) = q(k, l) \quad (9)$$

여기서 $P(H_0(k, l))$ 는 H_0 의 확률밀도함수를 $q(k, l)$ 는 사전음성부재확률을 나타낸다 [10]. 효율적인 사전음성부재확률 (a priori SAP)을 유도하기 위해 사전 SNR $\hat{\xi}(k, l)$ 방법을 적용하며, 재귀적 평균을 아래와 같이 표현한다 [9].

$$\zeta(k, l) = \beta\zeta(k, l-1) + (1-\beta)\hat{\xi}(k, l-1) \quad (10)$$

여기서 β 는 시간상수를 나타내며, 사전 SNR $\hat{\xi}(k, l)$ 는 다음과 같이 정의된다 [11].

$$\hat{\xi}(k, l) = a \frac{\hat{X}^2(k, l-1)}{\hat{\sigma}^2(k, l)} + (1-a) \max[\gamma(k, l) - 1, 0] \quad (11)$$

여기서 $a(0 < a < 1)$ 는 평활 파라미터를 나타내고, $\hat{X}^2(k, l-1)$ 는 이전 프레임의 추정된 음성 전력스펙트럼이며, $\hat{\sigma}^2(k, l)$ 는 추정된 잡음 전력스펙트럼이고, 최대함수 \max 는 음수를 제거하기 위한 함수로 사용된다. 그리고 사후 SNR $\gamma(k, l)$ 은 아래와 같이 나타낸다 [10].

$$\gamma(k, l) = |Y(k, l)|^2 / \hat{\sigma}^2(k, l) \quad (12)$$

다음은 수식 (10)을 재귀적 평균값을 이용하여 각 주파수 색인에 따른 'local' 로컬과 'global' 글로벌 평균값을 아래와 같이 구한다 [9].

$$\zeta_\lambda(k, l) = \sum_{i=0}^{2w_\lambda} h_\lambda(i) \zeta(k-i, l) \quad (13)$$

여기서 아래첨자 λ 는 'local' 또는 'global'을 나타낼 수 있고 [10], $h_\lambda(i)$ 는 크기 $2w_\lambda + 1$ 인 정규화 된 윈도우함수이다 [10].

위 수식 (13)과 각 프레임 l^{th} 의 k^{th} 주파수 도메인 음성의 위도 사이의 관계를 두 가지 형태의 $P_{local}(k,l)$ 와 $P_{global}(k,l)$ 파라미터로 아래와 같이 나타낸다.

$$P_{\lambda}(k,l) = \begin{cases} 0 & \text{if } \zeta_{\lambda}(k,l) \leq \zeta_{min}, \\ 1 & \text{if } \zeta_{\lambda}(k,l) \geq \zeta_{max}, \\ \frac{\log(\zeta_{\lambda}(k,l) / \zeta_{min})}{\log(\zeta_{max} / \zeta_{min})} & \text{otherwise,} \end{cases} \quad (14)$$

여기서 ζ_{min} 과 ζ_{max} 는 실험 치에 의한 상수를 나타내며, 약한 음성신호는 유지하고 잡음은 제거하기 위해 최대화 된다.

세 번째 파라미터 $P_{frame}(l)$ 은 인접구간의 음성에너지 기반으로 [10] 음성부재구간에서 잡음제거의 수준을 향상시키기 위해 사용된다.

$$\begin{aligned} &\text{If } \zeta_{frame}(l) > \zeta_{min} \text{ then} \\ &\quad \text{If } \zeta_{frame}(l) > \zeta_{frame}(l-1) \text{ then} \\ &\quad \quad P_{frame}(l) = 1 \\ &\quad \quad \zeta_{peak}(l) = \min \left\{ \max [\zeta_{frame}(l), \zeta_{p, min}], \zeta_{p, max} \right\} \\ &\quad \text{Else} \\ &\quad \quad P_{frame}(l) = \mu(l) \\ &\text{Else} \\ &\quad P_{frame}(l) = 0 \end{aligned} \quad (15)$$

여기서 $\zeta_{frame}(l) = \text{mean}_{1 \leq k \leq M/2+1} \{\zeta(k,l)\}$ 는 주파수 도메인의 평균을 나타내며, $\zeta_{p, min}$ 와 $\zeta_{p, max}$ 는 관측된 상수로 전이 여부를 결정하고, $\mu(l)$ 은 음성구간에서 잡음구간으로 "Soft transition"으로 정의된다 [10].

$$\mu(l) = \begin{cases} 0, & \text{if } \zeta_{frame}(l) \leq \zeta_{peak}(l) \cdot \zeta_{min} \\ 1, & \text{if } \zeta_{frame}(l) \geq \zeta_{peak}(l) \cdot \zeta_{max} \\ \frac{\log(\zeta_{frame}(l) / \zeta_{peak}(l) / \zeta_{min})}{\log(\zeta_{max} / \zeta_{min})}, & \text{otherwise,} \end{cases} \quad (16)$$

여기서 ζ_{peak} 은 ζ_{frame} 의 제한된 피크 값을 나타낸다. ζ_{frame} 이 증가하는 경우 음성의 시작부분의 왜곡을 방지하고, 또한 ζ_{frame} 이 감소하는 경우 음성구간에서 비 음성 구간으로 전이지연으로 음성의 파워가 약한 부분의 왜곡을 예방할 수 있다고 가정한다 [10].

마지막으로, 음성부재확률을 아래와 같이 정의한다 [9].

$$\hat{q}(k,l) = 1 - P_{local}(k,l) \cdot P_{global}(k,l) \cdot P_{frame}(l). \quad (17)$$

결과적으로 만약 음성부재확률 $\hat{q}(k,l)$ 이전프레임과 인접한 주파수 성분보다 큰 값을 가지면 음성이 포함되지 않은 것으로 간주한다. 그러므로 음성부재확률 SAP가 1로 수렴하면, 음성존재확률은 0으로 수렴한다.

IV. 비선형 함수와 사전 음성부재확률을 이용한 최소통계법의 잡음전력편의 보상방법

본 장에서는 최소통계치의 잡음전력편의 보상을 위해 비선형 함수와 음성부재확률을 결합한 새로운 잡음전력편의 보상방법을 제안한다. 제안한 방법은 사후 SNR의 증감에 따라 비선형 함수인 시그모이드 함수를 적용하여 적응파라미터를 구성한다.

$$\delta(l) = \alpha \cdot \frac{1}{1 + \exp(-\beta \cdot SNR(l))} \quad (18)$$

여기서 $\delta(l)$ 은 수식 (19)의 사후 SNR로부터 유도된다. α 와 β 는 실험에 의한 경험상수로서 $\beta = 0.5$ 는 기울기를 결정하는 상수이며, $\alpha = 0.1$ 은 수식 (18)의 적응파라미터 $\delta(l)$ 의 최대값 δ_{max} ($\delta_{max} \ll 0.1$)와 최소값 $\delta_{min} = 0$ 을 결정하는 상수로 사용된다.

$$SNR(l) = 10 \cdot \log \left(\frac{\text{norm}(|Y(k,l)|^2, 2)}{\text{norm}(\hat{\sigma}_d^2(k,l-1), 2)} \right) \quad (19)$$

여기서 $\hat{\sigma}_d^2$ 는 추정된 잡음전력이며, norm 은 $\text{norm} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_K^2}$ 각 프레임의 주파수 성분벡터 "Euclidean" 길이를 나타낸다.

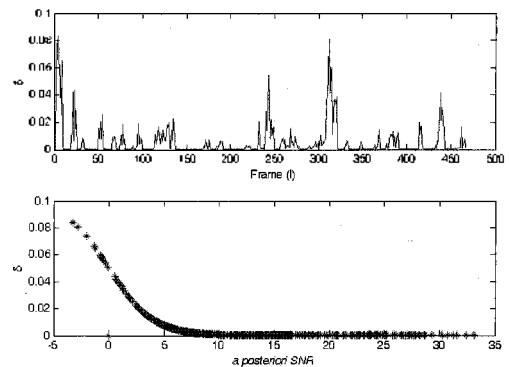


그림 2. 상단그림: 프레임 인덱스에서 적응파라미터. 하단그림: 사후 SNR에 기반 적응파라미터

Fig. 2. Top panel: plot of the adaptive parameter in the frame index. Bottom panel: adaptive parameter using a sigmoid function based on the a posteriori SNR.

그림 2는 잡음전력 보상 값을 위한 시그모이드 함수 기반 적응파라미터의 변화율을 나타낸다. 적응파라미터 $\delta(l)$ 는 사후 SNR이 증가할 경우 (즉, 음성구간) 0에 근접하는 값을 취하고, 반대의 경우 1에 근접한 값을 취한다. 결과적으로 잡음구간에서 취한 적응파라미터 값을 각 프레임의 주파수 성분의 음성부재확률 SAP 을 이용하여 아래와 같이 새로운 잡음전력 추정 값을 구할 수 있다.

$$\hat{\sigma}_d^2(k, l) = \hat{\sigma}_d^2(k, l) + \delta(l) \cdot \hat{q}(k, l) \quad (20)$$

여기서 음성부재확률 $\hat{q}(k, l)$ 가 1로 근접할 때 적응파라미터 $\delta(l)$ 는 δ_{\max} ($\delta_{\max} \ll 0.1$)에 접근한다. 그러므로 적응파라미터와 음성부재확률을 혼용한 방법으로 비 음성구간 (잡음구간)의 효율적인 잡음예측이 가능하다.

수식 (18)에서 구한 적응보상파라미터를 스무딩된 전력 스펙트럼 $P(k, l)$ 와 수식 (20)의 새로운 잡음전력 추정 값 $\hat{\sigma}_d^2$ 적용해 최소자승오류 (MSE)를 최소화함으로써 검증할 수 있다.

$$E\{(P(k, l) - \hat{\sigma}_d^2(k, l))^2 | P(k, l-1)\} \quad (21)$$

수식 (20)을 수식 (21)에 적용하여 적응보상파라미터 $\delta(l)$ 에 대해 정리하면 아래 수식 (22)가 유도된다.

$$\delta(l) = P(k, l) - \hat{\sigma}_d^2(k, l) \quad (22)$$

결과적으로 적응보상파라미터는 스무딩 전력스펙트럼과 수식 (4)의 추정된 잡음전력스펙트럼의 차이 값을 확인할 수 있다.

아래의 수식은 새로운 잡음전력 추정 값을 스무딩 파라미터에 적용한 새로운 스무딩 파라미터를 나타낸다.

$$\hat{\alpha}_{\text{opt}}(k, l) = \frac{\alpha_{\max} \cdot \alpha_c(l)}{1 + (P(k, l-1) / \hat{\sigma}_d^2(k, l-1) - 1)} \quad (23)$$

여기서 $\alpha_c(l)$ 는 오류추적 파라미터로 사용된다 [2].

V. 실험 결과 및 고찰

본 논문에서 제안된 알고리즘의 성능평가를 위해 white, babble과 car noise 환경과 극심한 비정상 잡음환경을 설정하여 객관적 테스트를 수행하였다. 실험에 사용된 DB는 남성과 여성 각 2명이 발성한 30개의 음성 샘플

표 1. 제안된 방법의 Segmental SNR (dB)과 ISDM의 비교와 객관평가

Table 1. Objective evaluation and comparison of the proposed method segmental SNR (dB) and the Itakura-Saito Distortion Measure (ISDM).

noise	method	SNR (dB)			
		from 0 to 15			
white	MS+SS	4.55	8.81	12.85	16.26
	Proposed+SS	5.01	9.55	13.79	16.74
babble	MS+SS	1.25	5.74	10.63	14.74
	Proposed+SS	1.37	5.94	10.94	15.07
car	MS+SS	3.63	7.84	11.77	15.41
	Proposed+SS	3.90	8.23	12.22	15.70
		ISDM			
white	MS+SS	1.14	0.77	0.50	0.33
	Proposed+SS	1.07	0.67	0.40	0.26
babble	MS+SS	0.26	0.09	0.03	0.01
	Proposed+SS	0.21	0.08	0.03	0.03
car	MS+SS	0.39	0.11	0.07	0.02
	Proposed+SS	0.33	0.09	0.06	0.02

플을 NOIZEUS [12]에서 사용하였고, 음성신호는 8kHz로 샘플링 하고 (STFT)를 사용하여 50% 오버랩 된 Hamming 윈도우 256 샘플을 이용하였다. 표 1은 기존방법 및 제안한 방법의 segmental SNR과 Itakura Saito Distortion Measure (ISDM) [12]의 결과를 보여주고 있다. 표 1의 결과로부터 babble 15 dB의 음질왜곡을 제외하고 제안한 방법이 기존의 방법보다 잡음제거 및 음질왜곡에서 향상된 결과를 보여주었다.

그림 3의 하단 그림은 개선된 잡음 추정 값을 사용하여 극심한 비정상 환경에서 잡음레벨을 추정한 것을 보여준다. 잡음구간에서 제안한 방법이 기존방법 최소 통계치 (MS)과 비교해 잡음 레벨을 잘 추적하는 것을 볼 수 있다. 상단그림은 추정된 잡음레벨을 이용하여 주파수 차감법과 결합시켜 잡음을 제거한 후의 그림이다. 그림에서 제안한 방법을 이용했을 때 잡음제거가 잘 된 것을 볼 수 있다.

최소 통계치(MS) 방법은 특정구간에서의 최소값을 추적하는 특성을 가지고 잡음레벨을 결정하므로 추정된 잡음레벨은 실제잡음레벨에 비해 하향편의의 단점을 가진다 [7][12]. 본 논문에서 이런 특성의 단점을 개선한 적응파라미터와 음성부재확률을 혼용한 잡음전력편의 보상방법이 잡음구간에서 효율적으로 잡음전력을 추정하는 것을 보여주었다. 따라서 제안한 방법이 기존 최소 통계치 방법보다 잔여잡음을 제거성능은 우수하고 음질왜곡은 최소화하는 기존의 trade off 한 특성에서 벗어났다고 할 수 있다.

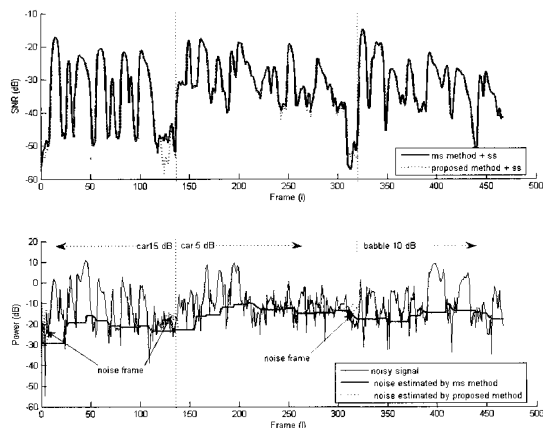


그림 3. 상단그림: 프레임 인덱스에서 적응파라미터. 하단그림: 사 후 SNR에 기반 적응파라미터
 Fig. 3. Top panel: plot of the adaptive parameter in the frame index. Bottom panel: adaptive parameter using a sigmoid function based on the *a posteriori* SNR.

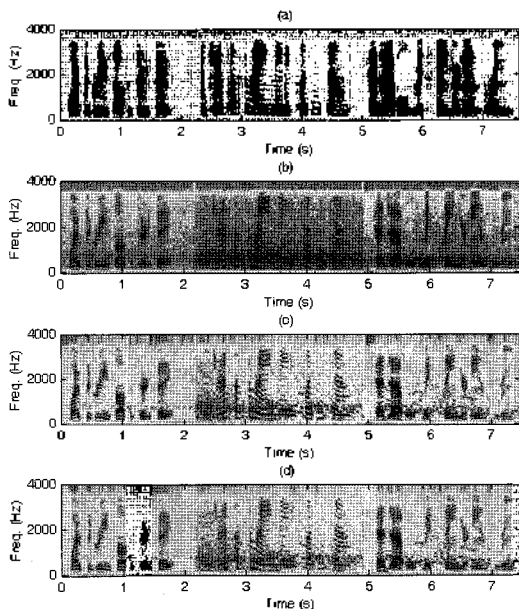


그림 4. Car noise 15 dB, car noise 5 dB, babble noise 10 dB의 비정상 잡음환경에서 음성강화 알고리즘의 시간 축 결과. (a) 음성신호, (b) 잡음에 오염된 음성신호, (c) MS+SS를 사용한 음성강화, (d) Propose+SS를 사용한 음성강화
 Fig. 4. Frequency domain results of speech enhancement for car noise at 15 dB, car noise at 5 dB, babble noise at 10 dB SNR in a nonstationary environments. (a) Original speech; (b) Noisy speech; (c) Enhancement speech using MS+SS; (d) Enhancement speech using Proposed+SS;

또한, 제안한 알고리즘의 음성향상성능을 평가하기 위해 남성, 여성의 상이한 문장을 다양한 잡음과 SNR (dB)에서 실험하였고, 잡음제거 후 음질향상의 성능을 평가하였다. 극심한 비정상 잡음환경의 실험을 위해 남성과

여성 그리고 각기 다른 특성과 SNR (dB)을 갖는 음성신호를 결합하여 실험환경을 설정하였다. 그림 4의 (c)에서 잡음 SNR이 car 15 (dB)에서 SNR car 5 (dB)로 변하는 2.2초 이후 잔여잡음을 볼 수 있고, (d)의 제안한 방법은 기존 최소 통계치 방법에 비해 비정상 잡음 환경에서도 잘 적응하여 잡음을 제거한 것을 확인 할 수 있다.

VI. 결론

본 논문에서는 최소 통계치 방법 (MS)이 잡음레벨이 변하는 실제 잡음환경에서 잡음전력추정 값이 실제 잡음 전력 값보다 하향 편이 되는 특성을 개선하기 위해 비정상 잡음환경에서도 효과적인 변형된 최소 통계치 알고리즘을 제안하였다. 제안된 방법은 비선형 함수를 적용한 적응보상파라미터와 사전 음성부재 확률 값을 혼용하는 잡음전력편의를 보상한방법이다. 새로운 방법을 다양한 잡음과 극심한 비정상 잡음에서 성능을 평가하여 기존 방법에 비해 우수한 결과를 나타내었다. 특히, 기존의 잔여잡음제거와 음질왜곡사이에 trade off 한 특성에서 벗어나 기존 최소 통계치 방법(MS)과 비교해 잡음제거 성능과 음질왜곡에서 향상된 결과를 나타내었다.

감사의 글

본 연구는 광운대학교 2008년도 교내연구비에 의해 수행되었습니다.

참고 문헌

1. 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.
2. R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Processing 9(5), 504-512, 2001.
3. I. Cohen, "Noise spectrum in adverse environments: improved minima controlled recursive averaging," IEEE Trans. Speech and Audio Proc, 11(5), 466-475, 2003.
4. L. Lin, W.H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," Electronics Letters, 39(9), 754-755, 2003.
5. R. Sundarajan, C.L. Philipos, "A noise-estimation algorithm for highly non-stationary environment," Speech Communication

- 48, 220–231, 2006.
6. 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," *J. Acoust. Soc. Kor.*, 27(1E), 1–10, 2008.
 7. 박운식, 장준혁, "강인한 음성향상을 위한 Minimum Statistics 와 Soft Decision의 확률적 결합의 새로운 잡음전력," *한국음향학회지* 26(4), 153–158, 2007.
 8. R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *SIGNAL PROCESSING* 86, 1215–1229, 2006.
 9. M. Berouti, M. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE int. Conf. Acoust. Speech Signal Processing*, 208–211, 1979.
 10. I. Cohen, "Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log–Spectral Amplitude Estimator," *IEEE Signal Processing Letters* 9(4), 113–116, 2002.
 11. Y. Ephraim and D. Malah, "Speech Enhancement using a minimum mean square error log–spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.* 23(2), 443–445, 1985.
 12. C. L. PHILIPPOS, *SPEECH ENHANCEMENT (Theory and Practice*, 1st edition, (CRC Press, Boca Raton, FL, 2007).

저자 약력

•이 수 정 (Soo-Jeong Lee)

한국음향학회지 제27권 제7호 참조

•이 강 성 (Gang-Seong Lee)



1986년 2월: 광운대학교 컴퓨터공학과 (공학사)
 1988년 8월: 광운대학교 컴퓨터공학과 (공학석사)
 1993년 2월: 광운대학교 컴퓨터공학과 (공학박사)
 2009년 2월: 서울대학교 음악대학원 작곡과 (예술학 석사)
 ※주관심분야: 음성/오디오 신호처리, 음성인식, 음악 예술공학, 미디어아트

•김 순 협 (Sun-Hyob Kim)

한국음향학회지 제27권 제7호 참조