논문 2012-49SP-3-11

서브밴드 가중치를 적용한 스펙트럼 최소값 추적을 이용하는 수정된 IMCRA 기반의 음성 향상 기법

(Speech Enhancement Based on Modified IMCRA Using Spectral Minima Tracking with Weighted Subband Selection)

박 윤 식*, 박 규 석*, 이 상 민**

(Yun-Sik Park, Gyuseok Park and Sangmin Lee)

요 약

본 논문에서는 잡음환경에서 음성 향상 (speech enhancement)을 위한 새로운 잡음전력 추정 방법을 제안한다. 음성 향상 알고리즘에 널리 적용되고 있는 IMCRA (improved minima controlled recursive averaging) 기법은 오염된 음성신호로부터 추정된 최소 전력 스펙트럼에 기반하여 잡음전력을 추정하는 기존의 방법을 개선하기 위해 간단한 음성 검출 알고리즘을 이용하여 대략적으로 음성 성분이 제거된 전력 스펙트럼에서 최소값을 추정함으로써 음성구간에서 발생할 수 있는 음성왜곡 문제점을 개선하였다. 하지만 비정상 잡음이나 신호 대 잡음 비 (SNR signal-to-noise ratio)가 낮은 환경에서는 음성 검출 성능이 저하되어 음성구간에서 음성왜곡이 발생되는 기존의 문제점이 여전히 발생된다. 따라서 제안된 방법에서는 향상된 잡음전력 추정을 위하여 기존의 IMCRA에서 추정된 최소 전력 스펙트럼에 대하여 스펙트럼 최소값 추적 (SMT, spectral minima tracking) 기법을 적용하고 IMCRA에 의한 최소값과 SMT에 의해 추정된 최소값을 서브밴드 (subband)에 따라 가중치를 적용하여 결합한다. 제안된 알고리즘은 기존의 방법과 주관적 및 객관적 음질평가 테스트를 통해 비교 평가한 결과 다양한 배경잡음 환경에서 향상된 성능을 보였다.

Abstract

In this paper, we propose a novel approach to noise power estimation for speech enhancement in noisy environments. The method based on IMCRA (improved minima controlled recursive averaging) which is widely used in speech enhancement utilizes a rough VAD (voice activity detection) algorithm which excludes speech components during speech periods in order to improves the performance of the noise power estimation by reducing the speech distortion caused by the conventional algorithm based on the minimum power spectrum derived from the noisy speech. However, since the VAD algorithm is not sufficient to distinguish speech from noise at non-stationary noise and low SNRs (signal-to-noise ratios), the speech distortion resulted from the minimum tracking during speech periods still remained. In the proposed method, minimum power estimate obtained by IMCRA is modified by SMT (spectral minima tracking) to reduce the speech distortion derived from the bias of the estimated minimum power. In addition, in order to effectively estimate minimum power by considering the distribution characteristic of the speech and noise spectrum, the presented method combines the minimum estimates provided by IMCRA and SMT depending on the weighting factor based on the subband. Performance of the proposed algorithm is evaluated by subjective and objective quality tests under various environments and better results compared with the conventional method are obtained.

Keywords : 음성 향상, 잡음전력 추정, 최소값 추적

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Department of Electronic Engineering, Inha University)
※ 본 연구는 서울시 산학연 협력사업(SS100022)의 지원에 의하여 이루어졌음.
접수일자: 2011년9월26일, 수정완료일: 2012년2월9일

Ⅰ. 서 론

현재 음성 향상 알고리즘은 잡음환경에서 음성 코덱이나 음성 인식기와 같은 음성신호 처리 시스템의 성능 및 통화 품질 향상을 위해 널리 적용되고 있다. 특히 음성 향상 알고리즘에서 추정된 잡음신호는 잡음제거 이득이나 주요 파라미터 추정에 직접적으로 영향을 미치기 때문에 향상된 음성 향상을 위해서는 다양한 잡음환경에서 효과적인 잡음신호 추정이 중요한 요소로 인식되고 있다[1]-[5]. 대략적으로 잡음신호를 추정하는 방법은 음성검출 (VAD, voice activity detection) 알고리즘을 이용하여 초반에 추정된 잡음신호를 비음성 구간에서만 갱신하여 잡음신호를 추정하는 방법과 잡음으로 오염된 음성의 전력 스펙트럼으로부터 추정된 최소값을 기반으로 잡음전력을 추정하는 방법으로 분류할 수 있다. 구체적으로 전력 스펙트럼의 최소값을 이용하는 잡음전력추정 기법은 음성신호와 잡음신호가 통계적으로 독립이라는 가정 하에 잡음이 부가된 음성신호의 전력 레벨이 빈번하게 잡음신호의 전력 레벨 까지 감소한다는 특성에 기반한 것으로써 갱신을 위한 적절한 크기의 윈도우 (window)를 사용하여 윈도우 구간 내에서 추정된 최소 전력에 바이어스 (bias) 보상을 통하여 잡음전력을 추정한다. 이와 관련하여 대표적인 방법으로 Martin에 의해서 제안된 MS (minimum statistics) 기반의 잡음 추정 방법은 정해진 윈도우에서 효과적인 최소값 추정을 위해서 최적환된 가변 스무딩 (smoothing) 파라미터를 적용하여 최소 전력 추정을 한다[3]. 가변 스무딩 매개 변수를 통해 구해진 최소 전력은 일반적으로 잡음전력의 평균보다 작게 추정되므로 분산을 기반으로 추정되는 바이어스 보상 함수을 통해서 보다 정확한 잡음전력을 추정한다. 그러나 MS 방식의 경우 윈도우 업데이트 구간이 음성구간에 존재할 경우 음성 성분에 의해 최소 전력이 크게 추정되고 이로 인해 음성 왜곡이 발생하며 최소값을 찾기위한 탐색 윈도우를 작게 하였을 경우, 무성음과 같이 작은 에너지를 갖는 음소가 감쇄 되는 단점을 지니고 있다. 따라서 이러한 최소 전력 추정에서 발생되는 문제점을 개선하기 위해 Cohen에 의해서 제안된 IMCRA (improved minima controlled recursive averaging, 향상된 최소값 제어 재귀 평균) 기법은 두 단계의 최소값 제어 재귀 평균 기법을 이용하여 추정된 최소 전력의 분산을 줄여주어 보다 정확한 잡음전력을 추정한다[5]. 구체적으로 IMCRA 방법에서는 기존의 최소값 제어 재귀 평균 기법을 적용하고 이에 기반한 주요 파라미터들로부터 도출된 VAD 알고리즘을 이용하여 입력된 신호에서 강인한 음성 성분을 제거하고 음성 성분이 제거된 신호로부터 정해진 윈도우 내에서 도출된 최소 전력을 기반으로 잡음전력이 추정된다. 하지만 IMCRA 추정법의 경우 비교적 적은 계산량으로 효과적인 잡음전력 추정이 가능하다는 장점이 있지만 정해진 윈도우보다 긴 시간동안 짧은 휴지기를 갖는 연속된 음성신호가 들어오게 되면 기존의 방법처럼 음성구간에서 신호의 최소 전력이 크게 추정되고 또한 비정상 잡음이나 SNR이 낮은 잡음환경에서는 적용된 VAD의 성능 저하로 효과적인 최소 전력의 추정이 어렵다는 문제점이 여전히 존재한다.

따라서 본 논문에서는 기존의 IMCRA 기법의 성능 향상을 위해서 IMCRA에 의해 추정된 최소 전력에 스펙트럼 편차를 기반으로 최소값을 추정하는 스펙트럼 최소값 추적 (SMT, spectral minima tracking) 기법을 적용하여 최소 전력을 추정한다 [1]. 또한 효과적인 최소 전력 추정을 위하여 음성과 잡음의 주파수 분포 특성을 고려하여 [8~10] IMCRA에 의해 도출된 최소 전력과 SMT에 의해 추정된 최소 전력을 서브밴드 (subband)에 따른 가중치로 결합하여 최종적으로 새로운 최소전력을 도출한다. 제안된 방법의 성능 평가를 위해 주관적인 음질평가 테스트인 MOS (mean opinion score) 테스트와 객관적인 음질평가 테스트 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. IMCRA 개요

Ⅱ장에서는 IMCRA 기반의 잡음전력 추정 기법에 대해서 간략하게 설명한다. 일반적으로 잡음전력을 추정하기 위해서 깨끗한 음성신호 $x(n)$과 잡음신호 $d(n)$이 상관성이 없다고 가정하여 잡음신호가 부가된 오염된 음성신호 $y(n)$을 $y(n)=x(n)+d(n)$이라고 하면 음성 존재 $H_1(k,l)$와 음성 부재 $H_0(k,l)$에 대한 가설을 다음과 같이 나타낼 수 있다.

$$H_0(k,l):\ Y(k,l)=D(k,l) \tag{1}$$

$$H_1(k,l):\ Y(k,l)=X(k,l)+D(k,l)$$

여기서 $X(k,l)$, $D(k,l)$는 각각 $x(n)$과 $d(n)$의 푸리에 변환에 (short-time Fourier transform, STFT) 의한 스펙트럼을 나타내고, k와 l은 주파수 인덱스 (index)와 프레임 인덱스를 나타낸다. 일반적으로 잡음전력 추정은 마이크로폰으로 입력된 오염된 음성신호 $Y(k,l)$에 대하여 음성부재 구간에서 입력신호의 시간축 스무딩을 적용하여 구해지며 음성 부재 구간에서 추정된 잡음신호의 분산을 $\overline{\lambda}_d(k,l) = E[|D(k,l)|^2]$라 하면, 다음과 같이 나타낼 수 있다[5].

$$H_0(k,l): \overline{\lambda}_d(k,l+1) = \alpha_d \overline{\lambda}_d(k,l) + (1-\alpha_d)|Y(k,l)|^2 \quad (2)$$
$$H_1(k,l): \overline{\lambda}_d(k,l+1) = \overline{\lambda}_d(k,l)$$

여기서 α_d는 시간축 스무딩 매개 변수를 의미하고 IMCRA 기법은 식 (2)에 음성부재 구간에서의 잡음전력 갱신을 위해 조건부 음성 존재 확률을 적용하여 다음과 같이 잡음전력을 추정한다.

$$\begin{aligned}\overline{\lambda}_d(k,l+1) &= \overline{\lambda}_d(k,l)p(k,l) \\ &\quad + [\alpha_d \overline{\lambda}_d(k,l) + (1-\alpha_d)|Y(k,l)|^2](1-p(k,l)) \\ &= \tilde{\alpha}_d(k,l)\overline{\lambda}_d(k,l) + [1-\tilde{\alpha}_d(k,l)]|Y(k,l)|^2\end{aligned} \quad (3)$$

여기서 $\tilde{\alpha}_d(k,l)$는 음성존재확률에 의해서 변화되는 스무딩 매개 변수로 $\tilde{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d)p(k,l)$로 정의되고, $p(k,l) = p(H_1(k,l)|\gamma(k,l))$는 *a posteriori* SNR의 조건부 음성 존재 확률로서 다음과 같이 나타낼 수 있다[5].

$$\begin{aligned}&p(k,l) \\ &= \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp\left(-\frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}\right)\right\}^{-1}\end{aligned} \quad (4)$$

여기서 $q(k,l)$는 음성 부재에 대한 사전 확률 $p(H_0)$ (*a priori* probability)을 나타내고, $\gamma(k,l)$, $\xi(k,l)$은 각각 *a posteriori* SNR과 *a priori* SNR로 다음과 같다.

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\lambda_d(k,l)} \quad (5)$$

$$\xi(k,l) = \frac{\lambda_s(k,l)}{\lambda_d(k,l)}. \quad (6)$$

IMCRA 방법은 식 (3)과 같이 $p(k,l)$에 의해서 잡음전력 추정을 위한 스무딩 매개 변수가 결정되기 때문에 $p(k,l)$은 잡음전력 추정에 주요한 요소로 작용한다. 따라서 $p(k,l)$을 도출하기 위한 식 (4)에서 $\psi = (1+\xi(k,l))\exp(-\gamma(k,l)\xi(k,l)/(1+\xi(k,l)))$라고 정의하면 $p(k,l)$은 다음과 같이 나타낼 수 있다.

$$p(k,l) = \frac{1-q(k,l)}{1+(\psi-1)q(k,l)} \quad (7)$$

식 (7)은 $q(k,l)$=1 또는 $q(k,l)$=0이 된다면 ψ 값에 상관없이 $p(k,l)$ 값이 결정된다는 것을 나타내고 있으며 결과적으로 음성신호의 대부분의 구간에서 $q(k,l)$는 0 또는 1의 값을 보이기 때문에 식 (3)에 의한 효과적인 잡음전력 추정을 위해서는 사전 음성 부재 확률 $q(k,l)$가 주요한 요소로 작용하는 것을 알 수 있다. 구체적으로 IMCRA 방법에서는 $q(k,l)$ 도출을 위하여 연속된 프레임에서의 음성의 강한 상관성을 고려하여 주파수 축과 시간 축에서 다음과 같이 스무딩을 적용한다[5].

$$S_f(k,l) = \sum_{i=-w}^{w} b(i)|Y(k-i,l)|^2 \quad (8)$$

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)S_f(k,l) \quad (9)$$

여기서 $b(i)$는 $2w$-1 길이의 해밍 (Hamming) 윈도우를 나타내고 w=1로 설정하였으며, $\alpha_s(0<\alpha_s<1)$는 시간축 에서의 스무딩 매개변수를 나타낸다. 스무딩된 신호 $S(k,l)$에 대하여 제한된 윈도우 길이 (0.8s~1.4s) 내에서 최소값 추정 방식에 의하여 최소값이 $S_{\min}(k,l)$ 구해지고[2], 이를 기반으로 기존의 전력의 최소값 추정에서 발생되는 문제점을 개선하기 위해 개략적인 VAD 적용을 위한 파라라미터 $\gamma_{\min}(k,l)$과 $\zeta(k,l)$이 다음과 같이 도출된다.

$$\gamma_{\min}(k,l) = \frac{|Y(k,l)|^2}{B_{\min}S_{\min}(k,l)} \quad (10)$$

$$\zeta(k,l) = \frac{S(k,l)}{B_{\min}S_{\min}(k,l)} \quad (11)$$

여기서 $B_{\min}$(=1.66)은 추정된 최소 잡음전력을 보상하기 위한 변수이고 도출된 $\gamma_{\min}(k,l)$과 $\zeta(k,l)$에 적절한 문턱값 γ_0과 ζ_0의 비교를 통하여 음성의 존재 유무를 판단하는 결정식 (decition rule)에 의한 VAD 알고리즘은 다음과 같이 표현된다[5].

$$I(k,l)=\begin{cases}1, & \text{if}(\gamma_{\min}<\gamma_0) \\ & \text{and}(\zeta(k,l)<\zeta_0) \quad (\text{speech is absent}) \\ 0, & \text{otherwis} \quad (\text{speech is present})\end{cases} \tag{12}$$

따라서 입력된 신호의 주파수 성분에서 최소전력을 추정할 때 음성 구간에서 발생되는 문제점을 개선하기 위해 식 (12)로 정의되는 VAD 알고리즘에 의해 강한 음성 부분을 제거하여 보다 효과적인 최소값 추정을 위한 입력신호 $Y(k,l)$는 $I(k,l)$를 기반으로 다음과 같이 수정된다[5].

$$\tilde{S}_f(k,l)=\begin{cases}\dfrac{\sum_{i=-w}^{w} b(i)I(k-i,l)|Y(k-1,l)|^2}{\sum_{i=-w}^{w} b(i)I(k-i,l)}, & \text{if } \sum_{i=-w}^{w} I(k-i,l)\neq 0 \\ \tilde{S}(k,l-1), & \text{otherwise}\end{cases} \tag{13}$$

$$\tilde{S}(k,l)=\alpha_s\tilde{S}(k,l-1)+(1-\alpha_s)\tilde{S}_f(k,l) \tag{14}$$

여기서 $\tilde{S}(k,l)$는 입력된 신호에서 강한 음성 성분이 제거된 후 시간-주파수축으로 스무딩 된 신호이며, $\tilde{S}(k,l)$를 기반으로 최소값 추정 방법을 통해서 최소전력 $\tilde{S}_{\min}(k,l)$이 추정된다. 추정된 최소값 $\tilde{S}_{\min}(k,l)$을 이용하여 $\widetilde{\gamma_{\min}}(k,l)$과 $\tilde{\zeta}(k,l)$이 다음과 같이 도출된다.

$$\tilde{\gamma}_{\min}(k,l)=\frac{|Y(k,l)|^2}{B_{\min}\tilde{S}_{\min}(k,l)} \tag{15}$$

$$\tilde{\zeta}(k,l)=\frac{S(k,l)}{B_{\min}\tilde{S}_{\min}(k,l)}. \tag{16}$$

최종적으로 추정된 파라미터 $\widetilde{\gamma_{\min}}(k,l)$과 $\tilde{\zeta}(k,l)$에 적절한 문턱값을 적용하여 잡음전력 추정을 위한 사전 음성 부재 확률 $q(k,l)$ 다음과 같이 표현된다[5].

$$\hat{q}(k,l)=\begin{cases}1, & \text{if } \tilde{\gamma}_{\min}(k,l)\le 1 \\ & \text{and } \tilde{\zeta}(k,l)<\zeta_0 \\ (\gamma_1-\tilde{\gamma}_{\min}(k,l)) & \text{if } 1<\tilde{\gamma}_{\min}(k,l)<\gamma_1 \\ /(\gamma_1-1), & \text{and } \tilde{\zeta}(k,l)<\zeta_0 \\ 0, & \text{otherwise}\end{cases} \tag{17}$$

Ⅲ. 제안된 서브밴드 가중치를 적용한 스펙트럼 최소값 추적 기반의 음성 향상 알고리즘

Ⅱ장에서는 IMCRA에 의한 잡음전력 추정에 대하여 설명하였다. Ⅲ장에서는 기존의 IMCRA 기법의 성능 향상을 위해서 새로운 음성향상 기법을 제안한다. 제안된 방법에서는 IMCRA에 의해 추정된 최소전력에 스펙트럼 최소값 추적 (SMT, spectral minima tracking) 기법을 적용하고 IMCRA에 의한 최소전력과 SMT에 의해 추정된 최소전력을 음성과 잡음의 주파수 분포 특성을 고려하여[8~10] 서브밴드에 따라 가중치를 주어 결합하여 새로운 최소전력을 도출한다. 구체적으로, 기존에 최소전력을 추정할 때 윈도우 갱신이 음성구간에서 발생하여 최소전력이 크게 추정되고 이로부터 음성왜곡이 발생되는 문제점을 개선하기 위해 제안된 알고리즘은 IMCRA에 의해 추정된 최소전력에 스펙트럼 편차를 기반으로 최소값을 추정하는 SMT 기법을 적용하여, 기존의 방법에 의해 추정된 최소전력의 분산을 감소시킴으로서 음성구간에서 발생할 수 있는 잡음 추정 오차와 음성왜곡을 감소시킨다[1]. 또한 제안된 방법에서는 보다 효과적인 최소전력 추정을 위해 주파수 영역에서 음성 정보의 대부분이 저주파 대역에 분포하는 음성신호의 주파수 분포 특성을 고려하여[8~10] IMCRA와 SMT에 의해 추정된 최소전력을 서브밴드에 따라 가중치를 주어 결합하여 새로운 최소전력을 도출한다. 즉, sigmoid 함수에 의한 도출된 가중치를[6] 서브밴드 별로 적용하여 음성의 주파수 성분이 주로 분포하고 있는 저주파 대역에서는 SMT에 의해 분산이 감소된 최소전력에 가중치를 주고 잡음신호 성분이 주로 분포하고 있는 고주파 대역에서는 기존의 IMCRA에 의해 추정된 최소전력에 가중치를 주어 서브밴드에 따라 선별적으로 결합하여 새로운 최소전력을 도출한다.

제안된 방법에서는 주파수 영역에서의 서브밴드 적용을 위해 다음과 같이 주파수 성분에 대하여 서브밴드를 적용한 신호 $Y_{sb}(m,l)$를 도출한다[7].

$$Y_{sb}(m,l)=\frac{\sum_{n=f_L(l)}^{f_H(l)}|Y(k,l)|^2}{f_H(m)-f_L(m)} \tag{18}$$

여기서 m $(1\le m\le N_{sb})$은 서브밴드 인덱스를 의미하여 본 논문에서는 128-point FFT (fast Fourier

transform)에 의해 도출된 64개의 주파수 성분을 16 ($N_{sb}=16$)개의 서브밴드로 구성하였다. 또한 $f_H(m)$와 $f_L(m)$는 m번째 서브밴드의 결합을 위한 low 밴드 high 밴드의 주파수 결합 table로써 다음과 같다[7].

$$f_L=\{2,4,6,8,10,12,14,17,20,23,27,31,36,42,49,56\},$$
$$f_H=\{3,5,7,9,11,13,16,19,22,26,30,35,41,48,55,63\} \tag{19}$$

따라서 식 (18)의 서브밴드를 IMCRA 알고리즘에 적용하여 식 (1)은 다음과 같이 나타낼 수 있다.

$$H_0(m,l):\ Y(m,l)=D(m,l) \tag{20}$$

$$H_1(m,l):\ Y(m,l)=X(m,l)+D(m,l) \tag{21}$$

적용된 서브밴드를 기반으로 IMCRA 알고리즘의 식 (14)로부터 도출된 최소전력을 $\tilde{S}_{\min}(m,l)$라 하면 음성 구간에서 최소전력의 분산을 줄이기 위해 $\tilde{S}_{\min}(m,l)$에 대하여 SMT를 적용하여 추정된 최소전력 $P_{\min}(m,l)$은 다음과 같이 표현된다[1].

$$P_{\min}(m,l)=\begin{cases}\gamma P_{\min}(m,l-1)+\dfrac{1-\gamma}{1-\beta}(\tilde{S}_{\min}(m,l)-\beta\tilde{S}_{\min}(m,l-1)), & \\ \qquad\qquad \text{if } P_{\min}(m,l-1)<\tilde{S}_{\min}(m,l) & \\ \tilde{S}_{\min}(m,l), \qquad\qquad \text{otherwise} & \end{cases} \tag{22}$$

여기서 γ=0.998, β=0.96으로 설정되었으며[1], SMT에 의해 추정된 최소전력 $P_{\min}(m,l)$은 서브밴드에 따른 가중치 따라 IMCRA에 최소전력 $\tilde{S}_{\min}(m,l)$과 다음과 같이 결합된다.

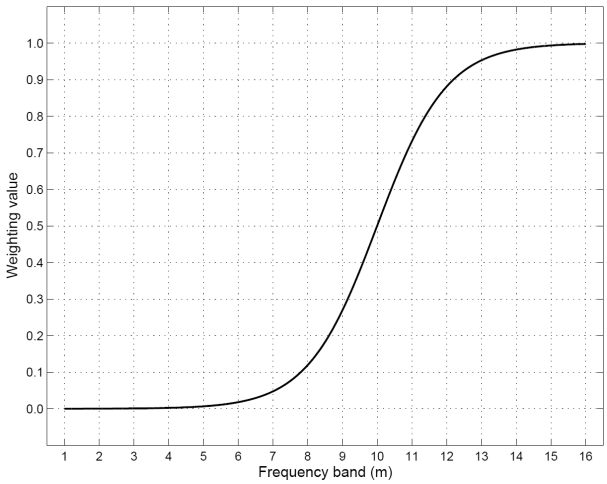

그림 1. Sigmoid 함수에 의한 서브밴드에 따른 가중치
Fig. 1. Weighting value by sigmoid function in subband.

$$\Gamma_{\min}(m,l)=\alpha_w(m)\tilde{S}_{\min}(m,l)+(1-\alpha_w(m))P_{\min}(m,l) \tag{23}$$

여기서 $\Gamma_{\min}(m,l)$는 제안된 최소전력을 의미하며 $\alpha_w(m)$는 sigmoid 함수에 의해 0~1의 값으로 도출되는 가중치로서, 적용된 sigmoid 함수에 다음과 같이 표현된다 [6].

$$\alpha_w(m)=\frac{f\exp[\beta(\mathrm{m}+\mathrm{s})]}{1+\exp[\beta(\mathrm{m}+\mathrm{s})]}+\sigma \tag{24}$$

식 (24)에서 파라미터들은 테스트 DB에 대한 성능의 최적화를 고려하여 f=1, β=1, s=-10, σ=0으로 설정하였다. 그림 1은 식 (24)의 sigmoid 함수에 의해 서브밴드에 따라 도출되는 가중치를 보여주고 있다. 그림 1로부터 음성 성분이 주로 분포 되어 있는 저주파의 서브밴드에서는 가중치 $\alpha_w(m)$=0이 적용되어 SMT에 의해 추정된 최소전력 $P_{\min}(m,l)$이 선택되어 지고 잡음신호 성분이 많은 고주파 서브밴드에서는 가중치가 $\alpha_w(m)$=1로 적용되어 IMCRA에 추정된 최소전력 $\tilde{S}_{\min}(m,l)$이 선택되는 것을 볼 수 있다. 또한 그 외의 서브밴드에서는 0과 1사이의 가중치 값이 적용되어 SMT와 IMCRA에 의한 최소전력을 선별적으로 결합하는 것을 알 수 있다[8~10]. 따라서 IMCRA 기반의 잡음전력 추정에 주요한 요소인 식 (17)의 사전 음성 부재 확률 $q(m,l)$을 도출하는데 적용되는 식 (15)와 식 (16)의 $\tilde{\gamma}_{\min}(m,l)$과 $\tilde{\zeta}(m,l)$이 제안된 최소전력 $\Gamma_{\min}(m,l)$을 기반으로 다음과 같이 나타낼 수 있다.

$$\hat{\gamma}_{\min}(m,l)=\frac{|Y(m,l)|^2}{B_{\min}\Gamma_{\min}(m,l)} \tag{25}$$

$$\hat{\zeta}(m,l)=\frac{S(m,l)}{B_{\min}\Gamma_{\min}(m,l)}. \tag{26}$$

그림 2는 NOISEX-92 데이터베이스의 white 잡음이 SNR=0 dB로 부가된 오염된 음성신호에 대하여 음성신호 주파수 성분이 주로 분포되어 있는 서브밴드 m=2에서의 실제 잡음전력과 각각의 알고리즘에 의해 추정된 잡음전력 비교를 보여주고 있다. 구체적으로 그림 2(c)는 기존의 IMCRA에서 잡음전력의 갱신을 위한 선행 음성 부재 확률 $\hat{q}(m,l)$을 추정하는데 주요한 파라

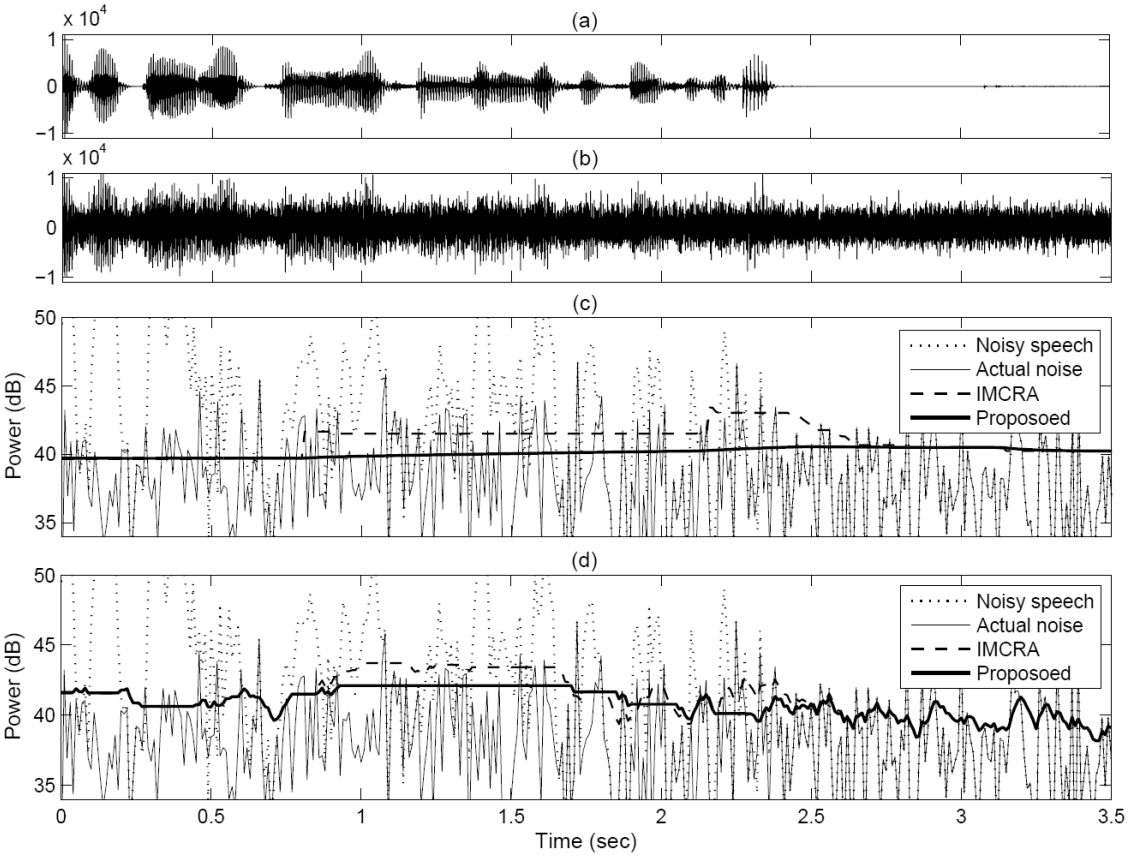

그림 2. 추정된 잡음전력 비교 (white noise, SNR=0 dB)
(a) 깨끗한 음성신호 파형 (b) 오염된 음성신호 파 (c) 사전 음성 부재 확률 $\hat{q}(m,l)$을 위해 추정된 기존의 $B_{\min}\tilde{S}_{\min}(m,l)$와 제안된 $B_{\min}\Gamma_{\min}(m,l)$ (d) 추정된 잡음전력.

Fig. 2. Comparison of estimated noise power (white noise, SNR=0 dB) (a) Clean speech waveform (b) Noisy speech waveform (c) the conventional $B_{\min}\tilde{S}_{\min}(m,l)$ and proposed $B_{\min}\Gamma_{\min}(m,l)$ for estimating *a priori* speech absence probability $\hat{q}(m,l)$ (d) the estimated noise power.

미터로 적용되는 기존의 최소전력에 기반하여 추정된 잡음전력 $B_{\min}\tilde{S}_{\min}(m,l)$과 제안된 알고리즘에 의해 도출된 잡음전력 $B_{\min}\Gamma_{\min}(m,l)$에 대한 실제 잡음전력과의 비교를 보여주고 있다. 그림 2(c)에서 점선으로 표기된 IMCRA에 의해 추정된 잡음전력 $B_{\min}\tilde{S}_{\min}(m,l)$으로부터 음성 존재하는 0.8~2.4초의 음성구간에서 기존의 방법에 의해 추정된 최소전력 기반의 잡음전력이 음성신호로 인해 실제 잡음전력보다 크게 추정되는 것을 볼 수 있다. 이러한 현상은 $B_{\min}\tilde{S}_{\min}(m,l)$가 주요하게 작용하는 $\hat{q}(m,l)$에 영향을 주어 최종적으로 $\hat{q}(m,l)$의 갱신을 통해 추정된 잡음전력 $\bar{\lambda}_d(m,l)$이 음성구간에서 크게 추정되고 이로 인해 음성 왜곡이 심해지는 문제점을 발생시킨다. 그림 2(d)에서 기존의 IMCRA에 추정된 최종 잡음전력이 실제 잡음전력보다 크게 추정된 것을 볼 수 있으며 이로부터 크게 추정된 잡음전력으로 인해 음성신호에 왜곡이 발생되는 것을 알 수 있다. 하지만 그림 2(c)에서 제안된 방법에 의해 추정된 최소전력 기반의 잡음전력 $B_{\min}\Gamma_{\min}(m,l)$은 $B_{\min}\tilde{S}_{\min}(m,l)$의 값을 유지하면서 최소전력이 커지는 음성구간에서는 분산을 감소시킴으로서 기존의 잡음전력 $B_{\min}\tilde{S}_{\min}(m,l)$ 보다 실제 잡음전력에 가깝게 추정하는 것을 볼 수 있으며, 따라서 그림 2(d)로부터 제안된 방법이 음성구간에서 기존의 방법보다 음성왜곡을 감소시키면서 효과적으로 잡음전력이 갱신되는 것을 알 수 있으며 음성이 존재하지 않는 잡음구간에서는 기존의 IMCAR처럼 잡음전력을 갱신하는 것을 볼 수 있다.

Ⅳ. 실험 및 결과고찰

본 논문에서는 제안된 알고리즘의 성능 평가를 위해 다양한 잡음환경에서 주관적 음질평가인 MOS 테스트와 객관적인 음질평가인 PESQ (perceptual evaluation of speech quality) 및 다음과 같이 전반적인 음질 C_{ovl}로 표현되는 composite measure 테스트를 수행하였다[11].

$$C_{ovl} = 1.594 + 0.805 S_{PESQ} - 0.512 S_{LLR} - 0.007 S_{WSS} \quad (27)$$

여기서 S_{PESQ}는 PESQ를 의미하며, S_{LLR}, s_{WSS}는 각각 log-likelihood ratio (LLR)과 weighted spectral slope (WSS)를 나타낸다. 테스트 샘플을 위해 8명의 화자로부터 얻은 8kHz로 샘플링 된 90개의 깨끗한 음성신호 수집하고 이를 4가지형태의 잡음을 다양한 SNR로 부가하여 오염된 음성신호를 생성하였으며 부가된 잡음은 NOISEX-92 데이터베이스를 이용하여 stationary white Gaussian noise (WGN), non-stationary WGN, car interior noise, F16 cockpit noise를 SNR은 -5, 0, 5, 10 dB로 부가하였다. 또한 제안된 알고리즘의 성능평가를 위해 제안된 방법을 Cohen에 의해 제안된 음성향상 알고리즘에 적용하고 도출된 결과 음성신호에 대하여 평가하였다.

표 1은 다양한 잡음 및 SNR 환경에서의 MOS 테스트 결과를 보여주고 있다. MOS 테스트 결과는 10명의 청취자에 의해 수행된 MOS의 평균값으로 나타내었으며 표 1로 부터 적용된 잡음환경에 대하여 제안된 알고리즘이 기존의 방법보다 비교적 향상된 MOS를 보이는 것을 볼 수 있다. 또한 각각의 알고리즘에 의한 PESQ 수치를 보여주고 있는 표 2로부터 모든 잡음환경에 대하여 제안된 알고리즘이 기존의 알고리즘보다 향상된 PESQ 수치를 보여 주고 있는 것을 알 수 있으며, 최종적으로 composite measure에 대해 수치를 보여주고 있는 표 3로부터 제안된 알고리즘이 기존의 방법보다 모

표 1. 다양한 잡음환경에서 IMCRA와 제안된 방법에 의한 MOS 수치.
Table 1. MOS obtained by IMCRA and the proposed method under various noise environments.

SNR	Stationary WGN		Non-stationary WGN		Car interior noise		F16 cockpit noise	
(dB)	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed
-5	1.29	1.29	1.29	1.36	3.79	3.93	1.07	1.14
0	1.79	1.93	1.43	1.71	3.93	4.07	2.07	2.29
5	2.50	2.71	2.71	2.86	4.43	4.50	2.50	2.71
10	3.36	3.64	3.29	3.50	4.36	4.36	3.50	3.79

표 2. 다양한 잡음환경에서 IMCRA와 제안된 방법에 의한 PESQ 수치.
Table 2. PESQ scores obtained by IMCRA and the proposed method under various noise environments.

SNR	Stationary WGN		Non-stationary WGN		Car interior noise		F16 cockpit noise	
(dB)	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed
-5	1.331	1.340	1.340	1.398	2.768	2.785	1.425	1.444
0	1.534	1.550	1.408	1.477	3.047	3.060	1.868	1.891
5	1.963	1.977	1.956	1.991	3.346	3.352	2.220	2.238
10	2.320	2.327	2.333	2.353	3.638	3.642	2.586	2.595

표 3. 다양한 잡음환경에서 IMCRA와 제안된 방법에 의한 composite measure.
Table 3. composite measure obtained by IMCRA and the proposed method under various noise environments.

SNR	Stationary WGN		Non-stationary WGN		Car interior noise		F16 cockpit noise	
(dB)	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed	IMCRA	Proposed
-5	1.270	1.279	1.260	1.314	2.907	2.929	1.525	1.542
0	1.612	1.632	1.385	1.453	3.234	3.250	2.075	2.098
5	2.134	2.150	2.107	2.143	3.595	3.604	2.481	2.498
10	2.554	2.563	2.542	2.561	3.942	3.947	2.930	2.938

든 잡음환경에서 개선된 composite measure 결과를 보여 주고 있는 것을 볼 수 있다. 따라서 잡음신호가 음성신호에 비해 비교적 작은 SNR=10 dB의 잡음환경에서부터 잡음신호의 의해 상당히 오염된 환경인 SNR=-5 dB까지의 다양한 잡음환경에서 모든 잡음 종류에 대해 주관적 및 객관적 음질평가로 부터 제안된 알고리즘이 기존의 방법보다 향상된 음질평가 수치를 보임으로서 제안된 알고리즘이 기존의 방법보다 개선된 성능을 보인다는 것을 알 수 있다.

V. 결 론

본 논문에서는 잡음환경에서 효과적인 잡음전력 추정을 위하여 기존의 IMCRA에서 추정된 최소전력에 SMT 기법을 적용하고 IMCRA와 SMT에 의해 추정된 최소전력을 서브밴드에 따라 가중치를 적용하여 결합하는 새로운 음성향상 알고리즘을 제안하였다. 제안된 알고리즘은 다양한 잡음환경에서 주관적 및 객관적 음질평가 테스트로부터 기존의 방법보다 향상된 결과를 나타내었다.

참 고 문 헌

[1] G. Doblinger, “Computationally efficient speech enhancement by spectral minima tracking in subbands,” in *Proc. EUROSPEECH*, vol. 2, pp. 1513-1516, 1995.

[2] R. Martin, “Spectral subtraction based on minimum statistics,” in *Proc. Eur. Signal Processing Conf.*, pp. 1182-1185, 1994.

[3] R. Martin, “Noise power spectral density estimation based. on optimal smoothing and minimum statistics,” *IEEE. Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504 - 512, July 2001.

[4] I. Cohen, B. Berdugo, “Noise estimation by minima controlled recursive averaging for robust speech enhancement,” *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12-15, Jan. 2002.

[5] I. Cohen, “Noise spectrum estimation in adverse environments: Improved minima controller recursive averaging,” *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 466-475, Sep. 2003.

[6] Y. -S. Park and J. -H. Chang, “A novel approach to a robust a priori SNR estimator in speech enhancement,” *IEICE Trans. on Communications*, vol. E90-B, no.8, pp 2182-2185 Aug. 2007.

[7] TIA/EIA/IS-127, Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, 1996.

[8] G. D. Wu and C. T. Lin, “Word boundary detection with mel-scale frequency bank in noise environment,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 541 - 554, May 2000.

[9] Y. Tian, J. Wu, Z. Wang, and D. Lu, “Robust noisy speech recognition with adaptive frequency bank selection,” in *Proc. ICMI*, pp.75-80, 2002.

[10] B.F. Wu, K.C. Wang, “Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments,” *IEEE Trans. Speech Audio Process.* vol. 13, no. 5, pp. 762-775, Sept. 2005.

[11] Yi Hu and P. C. Loizou, “Evaluation of objective quality measures 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>

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>

박 규 석(학생회원)
2010년 2월 인하대학교 전자공학과 학사
2012년 2월 인하대학교 전자공학과 석사
2012년 3월～현재 인하대학교 전자공학과 박사과정
<주관심분야 : 음성 신호처리, 잡음제거, 심리음향>