논문 2012-49SP-1-13

전역 음성 부재 확률 기반의 향상된 최소값 제어 재귀평균기법을 이용한 음성 향상 기법

(Speech Enhancement Based on Improved Minima Controlled Recursive Averaging Incorporating GSAP)

송 지 현*, 방 동 혁*, 이 상 민**

(Ji-Hyun Song, Dong-Hyeouck Bang, Sangmin Lee)

요 약

본 논문에서는 향상된 최소값 제어 재귀 평균 기법 (improved minima controlled recursive averaging, IMCRA) 알고리즘의 잡음 전력 추정성능을 향상 시키기 위한 알고리즘을 제안한다. 기존의 IMCRA은 주파수 특성이 빠르게 변화하는 비정상적인 환경과 낮은 SNR을 갖는 상황에서 잡음 전력 추정에 직접적으로 영향을 미치는 음성 검출기의 성능이 강인하지 못한 단점이 있다. 본 연구에서는 강인한 음성 검출 성능을 위해서 기존 IMCRA의 음성 검출기에 전역 음성 부재 확률을 적용한 음성 향상 기법을 제안한다. 제안된 알고리즘의 성능 평가는 음성의 perceptual evaluation of speech quality (PESQ)와 composite measure를 통한 음질을 평가하였다. 실험 결과 다양한 잡음 환경 (car, white, babble)에서 전역 음성 부재 확률을 적용한 IMCRA의 음성 향상 기법이 향상된 결과를 보여주었다. 특히, 비정상잡음 환경인 babble 5dB에서 PESQ 0.026, composite measure 0.029의 향상된 음질을 나타내었다.

Abstract

In this paper, we propose a novel method to improve the performance of the improved minima controlled recursive averaging (IMCRA). From an examination for various noise environment, it is shown that the IMCRA has a fundamental drawback for the noise power estimate at the offset region of continuity speech signals. Especially, it is difficult to obtain the robust estimates of the noise power in non-stationary noisy environments that is rapidly changed the spectral characteristics such as babble noise. To overcome the drawback, we apply the global speech absence probability (GSAP) conditioned on both *a priori* SNR and *a posteriori* SNR to the speech detection algorithm of IMCRA. With the performance criteria of the ITU-T P.862 perceptual evaluation of speech quality (PESQ) and a composite measure test, we show that the proposed algorithm yields better results compared to the conventional IMCRA-based scheme under various noise environments. In particular, in the case of babble 5 dB, the proposed method produced a remarkable improvement compared to the IMCRA (PESQ = 0.026, composite measure = 0.029).

Keywords : Improved minima controlled recursive averaging (IMCRA), Speech enhancement, Global speech absent probability (GSAP)

Ⅰ. 서 론

음성 코덱이나 음성 인식기와 같은 음성 신호 처리 시스템의 성능 및 음성 품질 향상을 위한 전처리로서 다양한 음성 향상 기술이 개발되고 있다[1~3]. 이러한 음성 향상 기술은 크게 잡음 신호에서 잡음 전력을 추정

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

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 IT 융합 고급인력과정 지원사업의 연구결과로 수행되었음(NIPA-2011-C6150-1102-0001).
접수일자: 2011년8월17일, 수정완료일: 2011년11월25일

하는 단계와 이를 기반으로 각 스펙트럼에 대한 적합한 이득을 추출하는 단계로 이루어져 있다. 여기서 잡음 전력 추정은 스펙트럼 이득을 추출하는데 직접적으로 영향을 미치기 때문에 우수한 음성 향상을 위해서 정확한 잡음 전력 추정은 매우 중요한 요소이다.

현재 대표적인 잡음 전력 추정 방법으로 최소값 추정 기법 (minimum statistics, MS)과 향상된 최소값 제어 재귀 평균 기법 (improved minima controlled recursive averaging, IMCRA)에 기반한 잡음 전력 추정법이 우수한 성능을 보여주는 것으로 알려져 있고, 두 방법은 음성 신호와 잡음 신호가 통계적으로 독립이라는 가정 하에 잡음이 부가된 신호의 전력 레벨이 자주 잡음 신호의 전력 레벨 까지 감소한다는 관찰을 기반으로 한다[4~7]. 즉, 적절한 크기의 윈도우 (winodow)를 사용한다면 전체 주파수 대역에 대해서 전력의 최소값을 이용하여 잡음 전력을 추정하는 것이 가능하다는 것이다.

Martin에 의해서 제안된 MS 잡음 추정 방법은 정해진 윈도우에서 최소값 추정을 위해서 먼저 1차 회귀 방법을 이용하여 입력된 신호를 스무딩 (smoothing) 한다[5]. 여기서 스무딩 매개 변수는 이전프레임의 전력과 추정된 잡음 전력에 의해서 구해지는 최적화된 변수로서, 현재 프레임이 음성일 경우 현재 프레임에 1에 가까운 가중치를 주고, 잡음일 경우 0.5에 가까운 가중치를 주어 보다 정확한 최소 전력값 추정을 한다. 가변 스무딩 매개 변수를 통해 구해진 최소 전력값은 일반적으로 잡음 전력의 평균보다 작게 추정되므로 바이어스 보상을 통해서 보다 정확한 잡음 전력을 추정한다. 그러나 MS 방식의 경우 기존의 잡음 전력 추정 방식에 비해서 두 배 정도의 분산을 가지며, 최소값 탐색 윈도우를 작게 하였을 경우, 무성음과 같이 작은 에너지를 갖는 음소가 감쇄 되는 단점을 지니고 있다.

이에 반해서, Cohen에 의해서 제안된 향상된 최소값 제어 재귀평균 기법은 두 단계의 최소값 제어 재귀 평균 기법을 이용해서, 큰 값의 1차 회귀 스무딩 변수를 사용 가능하게 하여, 추정된 최소값의 분산을 줄여주어 보다 정확한 잡음 전력을 추정하게 한다[8]. 구체적으로 첫 번째 단계에서 기존의 최소값 제어 재귀 평균 기법에서 사용된 음성 검출알고리즘을 이용하여 개략적인 VAD를 수행하고, 그 결과를 이용하여 입력된 신호에서 강인한 음성 부분을 제거해준다. 그 후 두 번째 단계에서 강인한 음성 부분이 제거된 신호를 이용하여, 정해진 윈도우 내에서 구해진 최소 전력값을 기반으로 현재 프레임의 사전 음성 부재 확률이 구해지고, 이는 a posteriori SNR 기반의 음성 존재 확률을 구하는데 적용된다. 마지막으로 음성 존재 확률에 의해서 가변되는 스무딩 매개 변수를 이용하여 이전의 추정된 잡음 전력 신호를 갱신하여 현재의 잡음 전력을 추정한다.

위에서 언급된 잡음 전력 추정법의 경우 비교적 적은 계산량으로 견실한 잡음 전력 추정이 가능하다는 장점이 있지만, 성능향상을 위한 연구가 활발히 진행되고 있다[9~10]. 특히, IMCRA 잡음 전력 추정 알고리즘의 경우 정해진 윈도우보다 긴 시간동안 짧은 휴지기를 갖는 연속된 음성신호가 들어오게 되면 신호의 최소값이 크게 추정 되고, 이는 입력된 전력과 추정된 최소 전력사이의 비를 통해서 음성의 유무를 판단하는 음성 검출 알고리즘의 성능저하를 유발한다.

본 논문에서는 IMCRA의 잡음 전력 추정에 직접적으로 영향을 미치는 잡음 섞인 신호의 국부 에너지와 주어진 윈도우에서의 최소값의 비를 이용하여 주파수 밴드의 음성 유무를 판단하는 기존의 MCRA 기반의 음성 검출알고리즘 성능을 향상시키기 위해서 통계적 모델 기반의 우수한 음성 검출 성능을 보여주는 전역 음성 부재 확률(global speech absence probability, GSAP)을 적용하여 향상된 음성 향상 기법을 도출하였다. 제안된 알고리즘의 객관적인 성능을 평가하기 위해서 객관적인 음질 평가 방법인 perceptual evaluation of speech quality (PESQ) 와 composite measure 테스트를 하였고, 실험 결과 다양한 잡음 환경에서 향상된 음질을 보여주었다.

Ⅱ. IMCRA (Improved minima controlled recursive averaging) 개요

이번 장에서는 잡음 전력을 추정하는 IMCRA 알고리즘에 대해서 알아본다. 일반적으로 잡음 전력을 추정하기 위해서 음성 신호 $x(n)$ 와 가산 잡음 신호 $d(n)$가 상관성이 없다고 가정하고, 관측된 신호 $y(n)$를 $y(n)=x(n)+d(n)$ 로 나타낸다. 이를 기반으로 음성의 존재 $H_1(k,l)$와 음성 부재$H_0(k,l)$에 대한 가설을 다음과 같이 나타낼 수 있다.

$$\begin{aligned} H_0(k,l)&: Y(k,l)=D(k,l) \\ H_1(k,l)&: Y(k,l)=X(k,l)+D(k,l) \end{aligned} \tag{1}$$

여기서 $X(k,l)$,$D(k,l)$는 잡음이 없는 음성 신호와

잡음 신호의 푸리에 변환을 (short-time Fourier transform, STFT) 이용하여 구한 스펙트럼을 나타내고, k와 l 은 주파수 인덱스와 프레임 인덱스를 나타낸다. 위의 가설을 기반으로 일반적인 잡음 전력 추정은 음성 신호의 부재 구간에서 관측된 신호의 시간축 스무딩을 적용하여 구해지고, 음성 부재 구간에서의 잡음 신호의 분산을 $\overline{\lambda}_d(k,l) = E[|D(k,l)|^2]$라 하면, 다음과 같이 나타낼 수 있다.

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

여기서 α는 시간축 스무딩 매개 변수를 나타낸다. IMCRA에서는 식(2)에 조건부 음성 존재 확률을 적용하여 다음과 같이 잡음 전력을 추정한다.

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

여기서 $\overline{\alpha}_d(k,l)$는 음성존재확률에 의해서 변화되는 스무딩 매개 변수로 $\overline{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d)p'(k,l)$로 정의 되고, $p'(k,l) = p(H_1(k,l)|\gamma(k,l))$는 a posteriori SNR의 조건부 음성 존재 확률을 나타낸다. 일반적으로 $p'(k,l)$의 값은 음성 왜곡의 발생을 줄여주기 위해서 음성 쪽으로 바이어스 되어있다. 따라서 식(3)의 결과는 실제 잡음 전력보다 낮은 값으로 추정되므로 바이어스 보상을 통해서 향상된 잡음 전력을 추정한다.

$$\hat{\lambda}_d(k,l+1) = \beta \bullet \overline{\lambda}_d(k,l+1) \quad (4)$$

a posteriori SNR의 조건부 음성 존재 확률은 다음과 같이 구해진다.

$$p'(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-\frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)})\right\}^{-1} \quad (5)$$

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

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\lambda_d(k,l)},\ \xi(k,l) = \frac{\lambda s(k,l)}{\lambda_d(k,l)} \quad (6)$$

IMCRA에서 잡음 전력 추정은 $p'(k,l)$에 의해서 업데이트 스무딩 매개 변수가 결정되기 때문에 $p'(k,l)$의 성능에 의해서 직접적으로 영향을 받는다. 식 (5)에서

$$\psi = (1+\xi(k,l))\exp(-\frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}) \quad (7)$$

라고 정의 하면 음성 존재 확률은 다음과 같이 나타낼 수 있다.

$$p'(k,l) = \frac{1-q(k,l)}{1+(\psi-1)q(k,l)} \quad (8)$$

식 (8)을 통해서 보면 만약 $q(k,l)$=1 또는 $q(k,l)$=0이 된다면 ψ 값에 상관없이 $p'(k,l)$ 값이 결정된다는 것을 알 수 있다.

그림 1은 5 dB SNR을 갖는 car 잡음에서의 $q(k,l)$ 값과 $p'(k,l)$을 나타내고 있다. 그림1을 통해서 볼 수 있는 것처럼 많은 부분의 경우 $q(k,l)$=1 또는 $q(k,l)$=0의 값을 갖는 것을 볼 수 있다. 즉, IMCRA의 잡음 전력 추정 성능은 $q(k,l)$의 성능과 밀접하게 연관되어 있다는 것을 알 수 있다.

$q(k,l)$은 음성 부재에 대한 사전 확률로 다음과 같은 과정을 통해서 구해진다. 먼저 입력된 시간축 신호를 DFT를 통해서 주파수 축으로 변환하고, 연속된 프레임에서의 음성의 강한 연계성을 고려하기 위해서 이를 주파수 축과 시간축 에서 다음과 같이 스무딩을 한다.

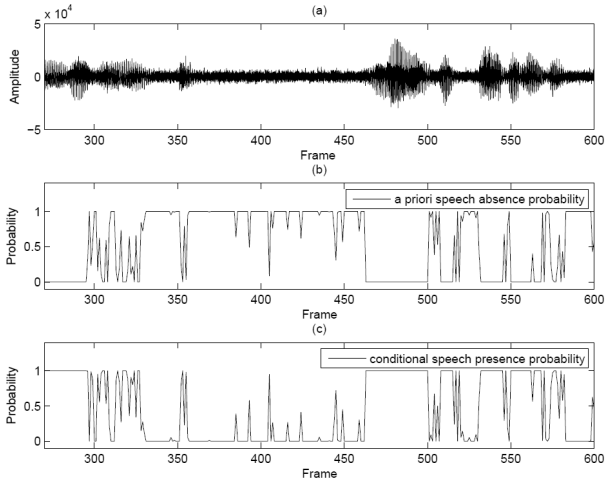


그림 1. car 잡음 (SNR =5 dB)에서의 사전 음성 부재 확률과 음성 존재 확률

Fig. 1. Probability under the car noise (SNR = 5dB) (a) Noisy speech waveform (b) A priori speech absence probability (c) Conditional speech presence probability.

$$S_f(k,l)=\sum_{i=-w}^{w} b(i)\,|Y(k-1,l)|^2$$
$$S(k,l)=\alpha_s S(k,l-1)+(1-\alpha_s)S_f(k,l) \quad (9)$$

여기서 $b(i)$ 는 2w+1 길이의 해밍윈도우를 나타내고, w 는 1의 값을 사용하였고, $\alpha_s\,(0<\alpha_s<1)$는 시간축 에서의 스무딩 매개변수를 나타낸다. 스무딩된 신호의 최소값을 검색하기 위해 정의된 윈도우 길이 (D 프레임) 내에서 스무딩된 입력 신호의 최소값이 구해지고, 이를 이용하여 각 주파수 밴드에 대해서 다음과 같이 개략적인 VAD가 수행된다 [8].

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

$$I(k,l)=\begin{cases}1 & (\gamma_{\min}<\gamma_0)\text{ and }(\zeta(k,l)<\zeta_0)\\ 0 & otherwise\end{cases} \quad (11)$$

여기서 $S_{\min}(k,l)$는 최소값 추정 방식에 의해서 구해진 최소값을 나타내고, $B_{\min}$은 구해진 최소 잡음 전력을 보상하기 위한 변수이다. $\gamma_{\min}(k,l)$, $\zeta(k,l)$ 와 정해진 문턱값 과의 비교를 통해서 개략적인 VAD가 수행된다.

두 번째 단계는 입력된 신호의 주파수 성분에서 강한 음성 부분을 제거하여 보다 안정적인 최소값 추정을 위해서 첫 번째 단계에서 구해진 VAD를 기반으로 다음과 같이 변화시킨다.

$$\widetilde{S}_f(k,l)=\begin{cases}\dfrac{\sum_{i=-w}^{w} b(i)I(k-i,l)|Y(k-1,l)|^2}{\sum_{i=-w}^{w} b(i)I(k-i,l)}, & \sum_{i=-w}^{w} I(k-i,l)\neq 0\\ \tilde{S}(k,l-1), & otherwise\end{cases} \quad (12)$$

$$\tilde{S}(k,l)=\alpha_s S(k,l-1)+(1-\alpha_s)\tilde{S}_f(k,l) \quad (13)$$

여기서 $\tilde{S}(k,l)$는 입력된 신호에서 강한 음성 성분이 제거된 후 시간-주파수축으로 스무딩 된 신호를 나타내고, 이 신호를 이용하여 최소값 추정 방법을 통해서 국부 에너지의 최소값이 추정된다 [8]. 구해진 국부 에너지의 최소값을 이용하여 $\tilde{\gamma}_{\min}(k,l)$, $\tilde{\zeta}(k,l)$가 식(10)과 유사하게 구해지고, 그 결과를 이용하여 사전 음성 부재 확률 $q(k,l)$이 구해진다.

$$q(k,l)=\begin{cases}1 & (\tilde{\gamma}_{\min}(k,l)\le 1)\\ & \text{and}(\tilde{\zeta}(k,l)<\zeta_0)\\ (\gamma_1-\tilde{\gamma}_{\min}(k,l))/(\gamma_1-1) & (1<\tilde{\gamma}_{\min}(k,l)\\ & \text{and}(\tilde{\zeta}(k,l)<\zeta_0)\\ 0 & otherwise\end{cases} \quad (14)$$

Ⅲ. Proposed IMCRA method based on GSAP

지금까지 Cohen에 의해서 제안된 IMCRA를 이용한 잡음 전력 추정에 대해서 알아보았다. 기존의 IMCRA 알고리즘의 경우 음성 존재 확률에 의해서 변화되는 잡음 전력 업데이트 매개 변수를 이용하여 향상된 잡음 전력 추정 결과를 보여준다. 음성의 강인한 성분을 제거해 주기위해서 사용하는 음성 검출 알고리즘의 경우 윈도우 내에서의 최소 전력과 입력 신호의 전력의 비를 통해서 이루어지기 때문에 잡음에서 음성으로 변화하는 구간을 빠르게 검출한다. 하지만, 윈도우 길이 보다 긴 짧은 휴지기간을 갖는 연속된 음성이 들어올 경우 최소값이 크게 추정되어, 음성의 꼬리 부분을 잡음 구간이라고 잘못 판단하게 된다. 이는 실제 잡음 전력보다 크

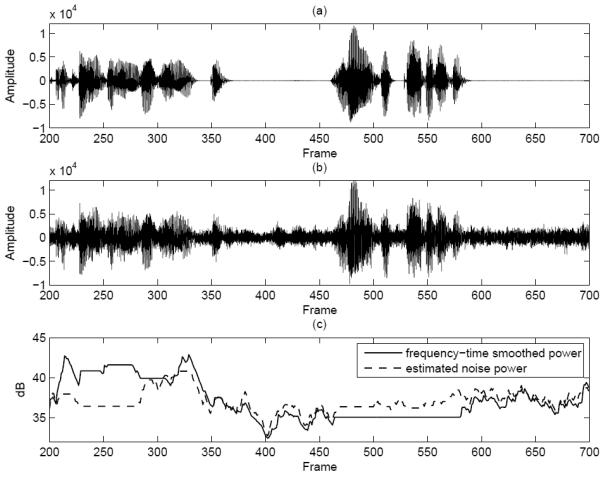


그림 2. babble 잡음 (SNR =5 dB)에서의 강인한 음성 성분이 제거된 전력(실선)과 IMCRA에서 최종적으로 추정된 잡음 전력(점선)

Fig. 2. Noise power under babble noise (SNR = 5 dB) (a) Clean speech waveform (b) Noisy speech waveform (c) Estimated power of Eq.(13) (bold line), estimated power of Eq.(4) (dashed line).

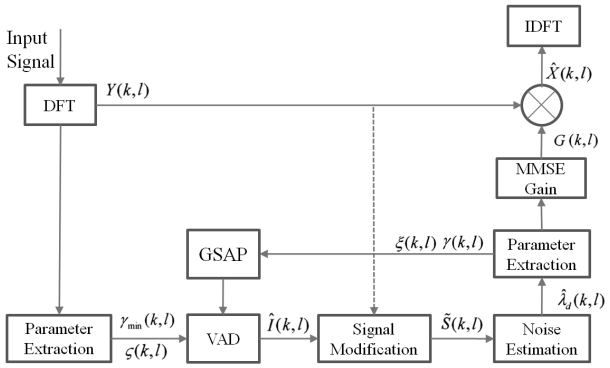


그림 3. 제안된 알고리즘의 전체 블록도
Fig. 3. Block diagram of the proposed method.

게 추정되게 되어 음성 신호를 왜곡하게 되어 음질 저하를 일으킨다.

그림 2는 babble 잡음 환경에서 5 dB SNR을 갖는 입력파일에 대해서 강한 음성 부분이 제거된 전력 (그림 (c)의 실선, 식(13)의 결과)와 최종적으로 구해진 잡음 전력(그림 (c)의 점선, 식(4)에 대한 결과)에 대한 그림을 나타낸다. 그림 2 (a)는 잡음이 없는 깨끗한 음성 파형을 나타내고, (b)는 babble 잡음이 첨가된 음성 파형을 나타낸다. 그림 2를 통해서 식(11)의 잘못 판단된 음성 검출 결과에 의해서 200~345 프레임에서 강한 음성 신호가 제거 되지 못하고, 그 결과로 인해서 280~345 프레임의 음성 꼬리 부분에서 잡음 전력이 크게 추정되는 것을 알 수 있다.

따라서 본 논문에서는 IMCRA에서 보다 안정적인 잡음 전력 추정을 위해서 음성 검출 알고리즘에 전역 음성 부재 확률을 적용한 향상된 IMCRA 알고리즘을 제안하고 그 과정은 그림 3에 나타내었다.

전역 음성 부재 확률의 경우 음성의 꼬리 부분에서 발생하는 검출 오류 문제점을 보완하는 계산 적으로 간단하고 효율적인 방법으로 알려진 강인한 음성 검출 알고리즘이다[11]. 전역 음성 부재 확률은 식(1)과 같은 음성 존재와 부재에 대한 가설을 기반으로 음성 신호와 잡음 신호의 전력이 평균이 0인 복소 가우시안 분포를 갖는다고 가정하면 두 가설을 조건으로 갖는 확률 밀도 함수는 다음과 같이 표현 된다.

$$p(Y(k,l)|H_0) = \frac{1}{\pi\lambda_d(k,l)} exp\left\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)}\right\}$$
$$p(Y(k,l)|H_1) = \frac{1}{\pi[\lambda_d(k,l)+\lambda_s(k,l)]} \cdot \exp\left\{-\frac{|Y(k,l)|^2}{\lambda_d(k,l)+\lambda_s(k,l)}\right\} \quad (15)$$

위의 두 가설을 이용하여 각 주파수 성분에 대한 음성 부재 확률은 다음과 같이 표현 된다.

$$p(H_0|Y(k,l)) = \frac{1}{1+p(H)/p(H_0)\Lambda(Y(k,l))} \quad (16)$$

여기서 $\Lambda(Y(k,l))$는 k 번째 주파수 대역의 우도비를 나타내고 다음과 같이 표현된다.

$$\Lambda(Y(k,l)) = p(Y(k,l)|H_1)/p(Y(k,l)|H_0) = \frac{1}{1+\xi(k,l)} exp\left[\frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}\right] \quad (17)$$

주파수 성분을 독립적이라고 가정하면 현재 프레임에서의 전역 음성 부재 확률은 다음과 같이 구해진다.

$$p(H_0|Y(l)) = \left[1+\epsilon\prod_{k=1}^{M}\Lambda(Y(k,l))\right]^{-1} \quad (18)$$

여기서 $\epsilon(=p(H_1)/p(H_0))$ 로 0.0625의 값을 주었고, M 은 주파수 밴드의 수를 나타낸다. 여기서 구해진 전역 음성 부재 확률을 IMCRA 잡음 추정 알고리즘에 적용하면 식(11)은 다음과 같이 나타낼 수 있다.

$$\hat{I}(k,l) = \begin{cases} 1 & (I(k,l)\equiv 1)\ \text{and}(p(H_0|Y(l) > TH_{GSAP}) \\ 0 & otherwise \end{cases} \quad (19)$$

전역 음성 부재 확률을 위와 같이 적용한 이유는 전역 음성 부재 확률의 경우 *a priori* SNR 추정시 Malah에 의해 제안된 musical 잡음을 효과적으로 줄여주는 Decision- Directed 추정 기법을 사용하기 때문에 음성 신호의 변화 구간에서 지연이 발생하고, 특히 잡음 신호에서 음성 신호로 변화되는 구간에서 지연에 의해서 음성 신호를 잡음 신호로 판단 될 수 있기 때문이다. 제안된 검출 룰은 기존의 IMCRA에 사용되는 음성 검출 알고리즘의 검출 룰에 의해서 잡음에서 음성으로의 변화 구간에서 빠르게 음성을 검출하고, 전역 음성 부재 확률에 의해서 음성의 꼬리 부분에서 잡음 신호로 잘못 검출되는 것을 방지하여 서로의 단점을 보완해 준다. 향상된 음성 신호를 구하기 위한 이득은 Ephraim -Malah에 의해서 제안된 최소 평균 평방 오차 (minimum mean square error, MMSE) 방식을 사용한다[12~13].

Ⅳ. 실험 및 결과

본 논문에서는 IMCRA의 음성향상 성능에 밀접하게 연관되어 있는 사전 음성 부재 확률의 향상된 추정을 위한 방법이 제안되었다. 구체적으로 개략적인 음성 검출 알고리즘에 전역 음성 부재확률에 대한 조건을 부여해 주어 잘못된 최소값 추정으로 인한 음성 검출 성능 저하를 줄여 주었다. 제안된 음성 향상 알고리즘의 성능을 평가하기 위해서 객관적 음질 평가 방식인 ITU-T P.862 PESQ와 composite measure를 이용하였다. 테스트를 위해서 남성, 여성에 의해서 각각 100개의 문장을 이용한 NTT 데이터베이스를 사용하였고, 잡음 환경을 부가하기 위해서 NOISEX-92 데이터베이스의 babble, car, white에서 5, 10, 15 dB 의 SNR을 갖는 테스트 파일을 만들었다. IMCRA에서 사용된 스무딩 매개변수는 $\alpha = 0.92, \alpha_d = 0.85, \alpha_s = 0.9, \gamma_0 = 4.6, \zeta_0 = 1.67$로 설정하였고, 최소값 추정을 위한 윈도우 길이는 $D = 120$, $U = 8$, $V = 15$ 로 설정하였다.

표 1은 제안된 방법과 기존 IMCRA 방법의 음질 성능 평가를 위해 실시한 PESQ 테스트의 성능을 나타낸다. 실험 결과, babble, car, white 잡음에 대해 평균적으로 각각 0.022, 0.014, 0.015 정도로 모든 잡음 환경에서 향상된 수치를 보여준다. 특히, babble 환경의 낮은 SNR에서 성능이 크게 향상되었는데 그 이유는 babble 잡음의 경우 음성에서의 스펙트럼과 근접한 주파수 대역에 에너지가 집중되어있기 때문에 식(9)에 의해서 음성 신호와 스무딩 되어 잡음 신호의 최소값이 증가하여 잘못된 최소값 추정이 이루어지기 때문이다.

표 2는 복수개의 기본적인 음질 평가 테스트로 구성되어 객관적인 음질을 평가하는 composite measure 테스트 결과를 보여주고, 다음과 같이 구성되어 있다.

표 1. 다양한 잡음 환경에서 기존 IMCRA와 제안된 방법의 PESQ 수치 비교

Table 1. Comparison of PESQ between the IMCRA and the proposed method.

Noise	Method	SNR (dB)		
		5	10	15
Babble	IMCRA	2.384	2.715	3.014
	Proposed	2.410	2.738	3.032
Car	IMCRA	3.641	3.888	4.088
	Proposed	3.653	3.903	4.104
White	IMCRA	2.145	2.486	2.839
	Proposed	2.165	2.499	2.852

표 2. 다양한 잡음 환경에서 기존 IMCRA와 제안된 방법의 composite measure 수치 비교

Table 2. Comparison of composite measure between the IMCRA and the proposed method.

Noise	Method	SNR (dB)		
		5	10	15
Babble	IMCRA	2.707	3.104	3.438
	Proposed	2.736	3.126	3.455
Car	IMCRA	3.928	4.206	4.436
	Proposed	3.939	4.221	4.452
White	IMCRA	2.336	2.726	3.114
	Proposed	2.356	2.740	3.125

$$C_{ovl} = 1.549 + 0.805 PESQ - 0.512 LLR - 0.007 WSS \quad (20)$$

여기서 로그 우도 비 (log-likelihood ratio, LLR)는 깨끗한 신호와 잡음 처리가 된 신호의 각각에 대해 추출된 LPC를 이용하여 복원된 신호의 차이를 로그 스케일로 측정하는 측정법을 나타내고, weighted-slope spectral distance (WSS) 는 정해진 프레임 내에서 인접한 주파수 밴드 사이의 관계의 왜곡도를 측정하는 측정법을 나타낸다[14]. composite measure에서 사용하는

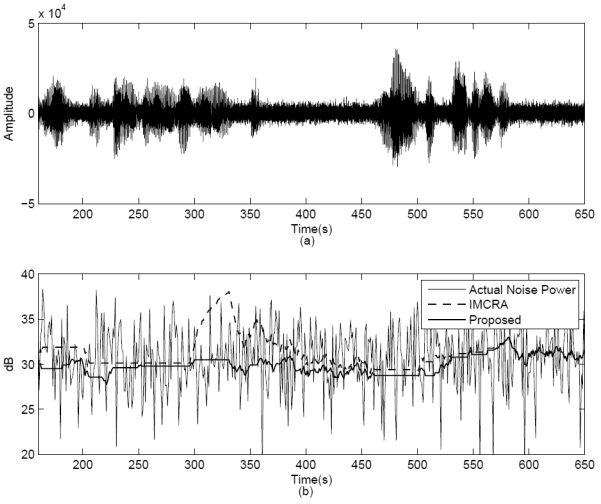


그림 4. white 잡음 (SNR = 5 dB) 파일에서 실제 잡음 전력과 IMCRA알고리즘과 제안된 알고리즘에 의해 추정된 잡음 전력 비교

Fig. 4. Comparison of estimated noise power under white noise (SNR = 5 dB) (a) noisy speech waveform (b) actual noise power (thin line), estimated noise power of IMCRA (bold line), estimated noise power of proposed (dashed line)

*PESQ*는 기존의 *PESQ*에서 음성의 왜곡과 잡음의 왜곡에 대한 측정치에 가중치를 더 주도록 수정된 측정법을 나타낸다. 표 2를 통해서 음성의 왜곡도와 인접한 프레임간의 왜곡도 측면에서도 제안된 알고리즘이 모든 잡음 환경에서 더욱 향상된 성능을 보여주는 것을 알 수 있다.

그림 4는 제안된 성능에 대해 보다 직관적으로 확인하기 위해서 실제 잡음 전력과 IMCRA와 제안된 알고리즘에 의해서 추정된 잡음 전력을 나타내고 있다. 그림 4 (a)는 5 dB SNR을 갖는 white 잡음 환경에서의 테스트 파일을 나타내고, 그림 4(b)의 실선, 점선, 굵은 실선은 각각 테스트 파일에 가산된 실제 잡음 전력, IMCRA를 통해서 추정된 잡음 전력, 제안된 알고리즘을 통해서 추정된 잡음 전력을 나타낸다. 그림4 (b)의 300~400 프레임 구간을 통해서 IMCRA의 경우 연속된 음성 신호에 의해서 잡음 전력의 추정이 잘못된 것을 알 수 있다. 이에 반해 제안된 방식의 경우 IMCRA 방식에 비해서 안정적인 잡음 전력 추정 결과를 보여준다.

V. 결 론

본 논문에서는 기존의 IMCRA에서 보다 향상된 잡음 전력 추정을 위해서 전역 음성 부재 확률을 적용한 음성 향상 알고리즘을 제안하였다. 구체적으로 IMCRA에서 입력된 신호에서 강인한 음성 부분을 제거하기 위해 사용되는 음성 검출 알고리즘에 통계적 모델 기반의 전역 음성 부재 확률을 적용하여 보다 안정적인 음성 검출 성능을 보여주었다. 특히, 전역 음성 부재 확률에 사용되는 특징 벡터의 경우 기존의 IMCRA에 사용되는 특징 벡터이기 때문에 적은 계산량의 추가로 향상된 잡음 전력 추정이 가능하게 했다. 제안된 음성 향상 기술의 성능을 평가하기 위해서 PESQ 와 composite measure 테스트를 하였고, 다양한 잡음 환경에서 제시된 전역 음성 부재 확률 기반의 향상된 IMCRA 기법이 기존의 IMCRA 보다 향상된 결과를 보여주었다.

참 고 문 헌

[1] S. F. Boll. “Suppression of acousitc noise in speech using spectral subtraction,” IEEE Transactions on Acoustics, Speech and Siganl Processing, ASSP-27(2), pp.113-120, Apr. 1979.

[2] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Transactions on Acoustics, Speech and Signal Processing, pp.113-120, Apr. 1979.

[3] I. Cohen and B. Berdugo, “Speech enhancement for non-stationary noise environment,” Signal Processing, pp.2403-2418, Nov. 2001.

[4] G. Doblinger, “Computationally efficient speech enhancement by spectral minima tracking in subbands,” Proc. 4th European Conf. Speech, Communication and Technology, EUROSPEECH’95, pp.1513-1516, Sep. 1995.

[5] R. Martin, “Spectral subtraction based on minimum statistics,” Proceeding of 7th EUSIPCO’94, Edinburgh, U.K., pp.1182-1185, Sep. 1994.

[6] I. Cohen and B. Berdugo, “Spectral enhancement by tracking speech presence probability in subbands,” Proc. IEEE Workshop on Hands Free Speech Communication, HSC’01, Kyoto, Japan, pp.95-98, Apr. 2001.

[7] I. Cohen and B. Berdugo, “ Noise estimation by minima controlled recursive averaging for robust speech enhancement,” IEEE Signal Processing Letters, pp.12-15, Jan. 2002

[8] I. Cohen, “Noise spectrum estimation in adverse environments : improved minima controlled recursive averaging,” IEEE Transactions on Speech and Audio Processing, pp.466-475, Sep. 2003.

[9] R. Martin, “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” IEEE Trans Acoustic, Speech and Audio Processing, pp.504-512, Jul. 2001

[10] S. Rangachari, P. C. Loizou and Y. Hu, “A noise estimation algorithm with rapid adaptation for highly nonstationary environments,” IEEE Conf. Acoustic, Speech Signal Processing, pp.305-308. May 2004.

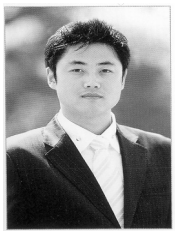
[11] N. S. Kim and J. H. Chang, “Spectral enhancement based on global soft decision,” IEEE Siganl Processing Letters, pp.108-110, May. 2000.

[12] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Transactions on Acoustics, Speech and Siganl Processing, ASSP-32(6), pp.1109-1121, Dec. 1984.

[13] Y. Ephraim and D. Malah, “Speech enhancement

using a minimum mean-square error log-spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Siganl Processing, ASSP-32(2), pp.443-445, Apr. 1985.
[14] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Audio, Speech and Language Processing, pp.229-238 Jan. 2008.

저 자 소 개



송 지 현(학생회원)
2007년 2월 인하대학교 전자공학과 학사
2009년 2월 인하대학교 전자공학부 석사
2009년 3월～현재 인하대학교 전자공학부 박사과정
<주관심분야 : 잡음제거, 음성검출, 음성 코딩>

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 책임연구원,
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>

방 동 혁(학생회원)
2007년 2월 광주대학교 전자공학과 학사
2009년 2월 인하대학교 전자공학부 석사
2011년 3월～현재 인하대학교 전자공학부 박사과정
1999년 6월～2003년 7월 에이스테크놀러지 사원
<주관심분야 : 잡음제거, 심리음향, 의용관련 응용 소프트웨어 구현>