

# 조건 사후 최대 확률 기반 최소값 제어 재귀평균기법을 이용한 음성향상

## Speech Enhancement Based on Minima Controlled Recursive Averaging Technique Incorporating Conditional MAP

김 중 모\*, 박 윤 식\*, 장 준 혁\*

(Jong-Mo Kum\*, Yun-Sik Park\*, Joon-Hyuk Chang\*)

\*인하대학교 전자공학부

(접수일자: 2008년 4월 16일; 수정일자: 5월 14일; 채택일자: 2008년 5월 14일)

본 논문에서는 기존의 최소값 제어 재귀 평균기법(minima controlled recursive averaging, MCRA) 알고리즘에 조건 사후 최대 확률(maximum a posteriori, MAP)을 적용한 음성향상을 제안한다. 기존의 MCRA는 파워스펙트럼에 평균을 취하고 각 서브밴드에서 음성 신호 존재 확률로 조절하는 스무딩 매개변수를 사용한다. 본 논문에서 제안된 알고리즘은 현재 프레임에 들어온 신호가 이전 프레임에서의 음성의 존재와 부재에 대한 조건을 부여해 주어 음성 신호 존재확률을 수정하여 음성향상에 적용한다. 제안된 음성 향상은 ITU-T P.862 perceptual evaluation of speech quality (PESQ)와 주관적 음질평가를 이용하여 평가하였고 기존의 MCRA 방법보다 향상된 결과를 나타내었다.

**핵심용어:** 최소값제어 재귀평균기법, 사후최대확률, 음성향상

**투고분야:** 음성처리 (2)

In this paper, we propose a novel approach to improve the performance of minima controlled recursive averaging (MCRA) which is based on the conditional maximum a posteriori criterion. A crucial component of a practical speech enhancement system is the estimation of the noise power spectrum. One state-of-the-art approach is the minima controlled recursive averaging (MCRA) technique. The noise estimate in the MCRA technique is obtained by averaging past spectral power values based on a smoothing parameter that is adjusted by the signal presence probability in frequency subbands. We improve the MCRA using the speech presence probability which is the a posteriori probability conditioned on both the current observation the speech presence or absence of the previous frame. With the performance criteria of the ITU-T P. 862 perceptual evaluation of speech quality (PESQ) and subjective evaluation of speech quality, we show that the proposed algorithm yields better results compared to the conventional MCRA-based scheme.

**Keywords:** minima controlled recursive averaging (MCRA), conditional maximum a posteriori (Conditional MAP)

**ASK subject classification:** Speech Signal Processing (2)

### I. 서론

실제적인 음성향상 시스템에서 잡음을 정확하게 추정하는 것이 가장 중요한 요소이며, 특히 비정상잡음 신호를 처리할 수 있어야 한다. 잡음 신호의 추정은 음성 향상 시스템의 품질에 미치는 영향이 크기 때문에 추정된 잡음 신호가 너무 작을 경우엔 자연스럽게 못한 잔류 잡음이

감지되며, 너무 클 경우엔 음성 신호가 둔탁하게 들리며 명료도가 떨어지게 된다. 전통적인 잡음 추정은 음성부재 구간에서 잡음의 평균을 구하는 것이 가장 일반적이며 [1]-[3], 이러한 음성 검출기 (voice activity detection, VAD)에 의존하는 추정은 조정하기가 어렵고 신호 대 잡음비가 낮은 응용분야에 사용될 때는 왜곡된 음성 신호를 출력하게 한다. 최근 진행 중인 연구들은 주로 Soft decision 방식을 적용하여 음성 영역에서도 잡음 신호의 파워 스펙트럼을 추정하고 [4], 대안으로 나온 최소 확률 잡음 신호 추정 (minimum statistics noise estimation,

MSNE) 기법은 VAD를 사용하지 않는 독창성을 가진다 [5]. 구체적으로 최적으로 신호의 파워 스펙트럼을 스무딩하고 최소 확률을 적용하는 것을 기본으로 하고, 파워 스펙트럼 스무딩 알고리즘은 시간과 주파수에 종속관계가 있는 스무딩 매개변수로 1차 회귀 시스템을 사용한다. 이때 스무딩 매개변수는 조건적인 평균 지층 오차를 최소로 하도록 최적화하여 비정상 잡음 신호를 추적한다. 그러나 이러한 잡음 추정은 특이점들에서만 민감하게 반응하고 전통적인 방법들보다 두 배 정도 큰 분산을 가진다. 더욱이 이 방법은 특히 최소 탐색 윈도우를 너무 작게 하면 작은 에너지를 갖는 음소를 더욱 약하게 만드는 단점이 있다 [8]. 계산량을 줄이고 효율적인 최소 추적 방법이 제안되기도 하였으나 [9], 잡음 신호의 에너지가 급상승하는 경우엔 잡음 추정이 느려지고 신호를 소멸시키는 문제점이 있다 [10].

한편 최근에 Cohen이 제안한 최소값제어 재귀평균기법 (minima controlled recursive averaging, MCRA) 은 서브밴드에서 신호 존재 확률로 조절하는 스무딩 매개변수를 이용하여 파워 스펙트럼에 평균을 취하는 방법으로 이러한 단점을 보완 한다 [6]. 각 서브밴드에서 신호의 존재는 잡음이 섞인 신호의 국부 에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한다. 이 비율과 특정 임계값과 비교하여 비율이 작으면 음성 신호가 없는 것으로 간주하는 방법을 사용하며, 음성 신호가 있는 부분과 없는 부분 사이에서 발생하는 변동을 줄이기 위해 시간 축으로도 평균을 취한다. 이것은 음성 신호의 존재에 있어서 프레임 사이에 높은 상관성이 있기 때문이다.

하지만 이러한 MCRA 알고리즘에도 몇 가지의 문제점들이 있다. 효율적인 지역 최소 추적 기술을 사용하여 계산의 복잡성은 줄였지만 갑작스러운 잡음 존재시 딜레이가 생기고, 또한 각 서브밴드에서 신호의 존재를 잡음이 섞인 신호의 국부에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한비율과 특정 한계 값만 가지고 비교를 하기 때문에 신뢰성이 떨어진다. 일반적으로, 음성의 활동은 인접한 프레임들과 강력한 상호 연관성이 있으므로, 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 가능성이 높다고 할 수 있고, 그 반대의 가능성도 맞다고 할 수 있다. 이러한 점에서 기존의 통계적 가정을 바탕으로 최근 연구에서는 hidden Markov model (HMM)을 이용한 행오버 알고리즘을 조건부 MAP (maximum a posteriori)에 적용하여 MCRA와 결합시켜 꼬리부분의 정보 왜곡을 최소화 시켰다 [7][11].

본 논문에서는 특정 임계값만을 사용하여 각 서브밴드

에서 음성신호의 유무를 추정하는 기존의 MCRA 방법대신 이전 서브밴드의 음성신호 유무에 대한 조건을 추가하여 더욱 신뢰성 있는 비교를 통해 음성 존재 확률의 성능을 향상시켜 우수한 잡음 추정 방법을 도출하여 향상된 MCRA를 제시 한다. 주관적인 음질 평가에서 제안된 음성 향상 방법이 기존의 MCRA 알고리즘 보다 좋은 성능을 나타내었으며 객관적 음질 평가 방법인 perceptual pvaluation of speech quality (PESQ)와 주관적 음질평가 테스트 결과에서도 향상된 성능을 보였다.

## II. MCRA (minima controlled recursive averaging) 개요

$x(n)$ 과  $d(n)$ 을 각각 음성신호와, 상관성이 없는 가산 잡음 신호라고 한다. 여기서  $n$ 은 이산시간을 나타낸다. 이때 관측되는 신호  $y(n)$ 은  $y(n) = x(n) + d(n)$ 으로 주어지고 부분적으로 중복으로 나누어 윈도우를 취한 다음 실시간 푸리에 변환 (short-time Fourier transform, STFT)을 이용하면 다음과 같이 나타낼 수 있다.

$$Y(k,l) = X(k,l) + D(k,l) \tag{1}$$

가 되는데 여기서  $Y(k,l)$ 은  $l$ 번째 프레임에서의  $k$ 번째 주파수 성분이 된다. 음성 향상 기법에서 사용되고 있는 기본가설은 음성의 부재와 존재 각각에 대해  $H_0(k,l)$ 과  $H_1(k,l)$ 라고 하면 다음과 같은 식으로 나타낼 수 있다.

$$\begin{aligned} H_0(k,l) &: Y(k,l) = D(k,l) \\ H_1(k,l) &: Y(k,l) = X(k,l) + D(k,l) \end{aligned} \tag{2}$$

여기서  $X(k,l)$ 과  $D(k,l)$ 은 각각 원래 음성 신호와 잡음 신호의 Fourier 변환 계수를 나타낸다. 여기서,  $\lambda_d(k,l) = E\{|D(k,l)|^2\}$ 이  $k$ 번째 서브밴드에서 잡음 신호의 분산이라고 하면, 추정하기 위해 음성 신호 부재 구간에서 관측된 신호에 시간의 반복 스무딩을 적용하면 다음과 같이 나타낼 수 있다.

$$\begin{aligned} H'_0(k,l) &: \hat{\lambda}_d(k,l+1) = \alpha_d \hat{\lambda}_d(k,l) + (1-\alpha_d) |Y(k,l)|^2 \\ H'_1(k,l) &: \hat{\lambda}_d(k,l+1) = \hat{\lambda}_d(k,l) \end{aligned} \tag{3}$$

여기서  $\alpha_d(0 < \alpha_d < 1)$ 는 스무딩 매개변수이다.  $H'_0, H'_1$

은 각각 가설에 근거하여 잡음 전력 갱신을 목적으로 한 음성 신호의 부재와 존재를 나타낸다. 음성 신호를 추정하는데 사용되는 식 (2)에서의 가설과 잡음 신호의 스펙트럼 갱신을 조절하는데 사용되는 식 (3)은 구별해야 한다. 즉, 음성신호가 존재 ( $H_1$ )할 때 음성신호의 부재 ( $H_0$ )라고 결정하는 것이 잡음 신호를 추정할 때보다 음성 신호를 추정할 때 더 위험스럽다. 그러므로 서로 다른 결정 법칙이 사용되고, 일반적으로  $H'$  보다는  $H_1$ 에 더 높은 신뢰를 두고 있으며,  $P(H_1|Y) \geq P(H'|Y)$ 인 것이다 [6].

$p(k,l) = P(H_1(k,l)|Y(k,l))$ 가 음성 존재의 조건 확률을 나타낸다고 하면, 식 (3)은 다음과 같다.

$$\begin{aligned} \hat{\lambda}_d(k,l+1) &= \hat{\lambda}_d(k,l)p(k,l) \\ &+ (\alpha_d \hat{\lambda}_d(k,l) + (1-\alpha_d)|Y(k,l)|^2)(1-p(k,l)) \\ &= \tilde{\alpha}_d(k,l)\hat{\lambda}_d(k,l) + (1-\tilde{\alpha}_d(k,l))|Y(k,l)|^2 \end{aligned} \quad (4)$$

여기서  $\tilde{\alpha}_d$ 는 음성 존재 확률로 조정하는 시변 스무딩 매개변수로서 다음과 같다.

$$\tilde{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d)p(k,l). \quad (5)$$

$p(k,l)$ 에 대한 추정은 다음과 같은 식을 사용 한다 [6].

$$\hat{p}(k,l) = \begin{cases} \alpha_p \hat{p}(k,l-1) + (1-\alpha_p) & \text{if } \frac{S(k,l)}{S_{\min}(k,l)} > \delta \\ \alpha_p \hat{p}(k,l-1) & \text{otherwise} \end{cases} \quad (6)$$

여기서  $\alpha_p$  ( $0 < \alpha_p < 1$ )는 스무딩 매개변수이고,  $\delta$ 는 음성신호 존재의 임계값이다. 특히,

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)|Y(k,l)|^2 \quad (7)$$

는 국부 에너지이고,  $\alpha_s$  ( $0 < \alpha_s < 1$ )는 매개변수이다. 국부 에너지의 최소값,  $S_{\min}(k,l)$ 과 임시변수,  $S_{tmp}(k,l)$ 은  $S_{\min}(k,0) = S(k,0)$  과  $S_{tmp}(k,0) = S(k,0)$  으로 초기화한다. 그리고 국부 에너지와 이전 프레임의 최소값을 비교하여 현재 프레임에 대한 최소값을 구한다.

$$S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\} \quad (8)$$

$$S_{tmp}(k,l) = \min\{S_{tmp}(k,l-1), S(k,l)\} \quad (9)$$

$L$ 개의 프레임을 모두 처리하고 나면 다음 식으로 초기

화 시킨다.

$$S_{\min}(k,l) = \min\{S_{tmp}(k,l-1), S(k,l)\} \quad (10)$$

$$S_{tmp}(k,l) = S(k,l) \quad (11)$$

식 (8)-(11)의 과정을 반복하여 최소값을 탐색한다.

$$S_r(k,l) \underset{H_0}{\overset{H_1}{>}} \delta \quad (12)$$

여기서  $S_r(k,l) = \frac{S(k,l)}{S_{\min}(k,l)}$ 이다.

### III. 제안된 조건 MAP (maximum a posterior)를 기반으로 한 향상된 MCRA

전통적인 MAP 기준은 아래와 같은 결정법을 따른다.

$$\frac{P(H_n = H_1|Y)}{P(H_n = H_0|Y)} \underset{H_0}{\overset{H_1}{>}} 1 \quad (13)$$

여기서  $H_n$ 은  $n$ 번째 프레임의 정확한 가정을 나타낸다. Bayes 법칙에 따르면 위의 기준을 아래와 같이 바꾸어 줄 수 있다.

$$\frac{p(Y|H_n = H_1)}{p(Y|H_n = H_0)} \underset{H_0}{\overset{H_1}{>}} \frac{P(H_n = H_0)}{P(H_n = H_1)} \quad (14)$$

바이어스된 것을 보상해 주기 위해 위의 식을 변형시켜 주면 다음과 같이 나타낼 수 있다.

$$\frac{p(Y|H_n = H_1)}{p(Y|H_n = H_0)} \underset{H_0}{\overset{H_1}{>}} \alpha \frac{P(H_n = H_0)}{P(H_n = H_1)} \quad (15)$$

여기서  $\alpha \geq 1$ 이다.

본 논문에서는 MCRA의 하나의 임계값만을 사용하여 각 서브밴드에서 신호의 존재를 추정할 때 이전 서브밴드의 신호가 음성을 포함하고 있는지 아닌지에 대한 조건을 추가하여 이 조건에 의해 두 가지의 임계값을 두어 더욱 신뢰성 있는 비교를 통해 신호 존재 확률의 성능을 더욱 좋게 해 주었다. 다음은 이전 서브밴드의 신호가 음성을

포함하고 있는지 아닌지에 대한 조건을 추가한 새로운 음성 신호 존재 확률의 정의를 유도하는 과정이다.

위의 식 (15)를 베이스의 불에 의해 바꾸어 표현하면 다음과 같다.

$$\frac{p(S_i|H_1)}{p(S_i|H_0)} \underset{H_0}{\overset{H_1}{>}} \alpha \frac{P(H_0)}{P(H_1)} \quad (16)$$

일반적으로, 음성의 활동은 인접한 프레임들과 강력한 상호 연관성이 있다. 즉, 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 가능성이 높다고 할 수 있고, 그 반대의 가능성도 맞다고 할 수 있다. 이러한 상호 연관성은 HMM을 이용한 행오버를 사용함으로써 통계 모델을 기반으로 한 VAD의 에러를 효과적으로 줄일 수 있다. 음성 활동 전후프레임들의 강력한 상호 연관성의 성질들로 인해 기존의 posterior probability  $P(H_n|X)$  대신 현재 프레임의 관찰결과와 이전 프레임의 결정을 조건으로 한 a posteriori probability  $P(H_n|X, H_{n-1})$  를 계산해 낼 수 있으므로 위 (16)식에 이전프레임의 조건을 추가 하여 주교 Bayes 법칙에 의해 아래와 같이 바꾸어 줄 수 있다.

$$\frac{p(H_n = H_1|S_i, H_{n-1} = H_1)}{p(H_n = H_0|S_i, H_{n-1} = H_1)} \underset{H_0}{\overset{H_1}{>}} \alpha \quad (17)$$

위의 식은 다시 아래와 같이 바꾸어줄 수 있다.

$$\frac{p(S_i|H_n = H_1, H_{n-1} = H_1)}{p(S_i|H_n = H_0, H_{n-1} = H_1)} \underset{H_0}{\overset{H_1}{>}} \alpha \frac{P(H_n = H_0|H_{n-1} = H_1)}{P(H_n = H_1|H_{n-1} = H_1)} \quad (18)$$

비록 현재 프레임의 음성 활동이 이전 프레임에 의존할 지라도, 현재 프레임의 음성 활동은 현재 프레임에서 관찰된 잡음 섞인 음성 신호의 DFT 계수의 분포에 지배적인 영향을 받는다. 더욱이,  $p(S_i|H_n = H_1, H_{n-1} = H_0)$ 와  $p(S_i|H_n = H_0, H_{n-1} = H_1)$ 에 분포된 매개변수들은 샘플 데이터의 부족으로 신뢰성 있는 추정이 불가하다. 이러한 이유로, 우리는 다음과 같이 간단한 가정을 할 수 있다.

$$p(S_i|H_n = H_j, H_{n-1} = H_i) = p(S_i|H_n = H_j), \quad i=0,1 \quad j=0 \quad (19)$$

그러면 위의 식 (18)을 다음과 같이 간략화 할 수 있다.

$$\frac{p(S_i|H_n = H_1)}{p(S_i|H_n = H_0)} \underset{H_0}{\overset{H_1}{>}} \alpha \frac{P(H_n = H_0|H_{n-1} = H_1)}{P(H_n = H_1|H_{n-1} = H_1)}, \quad i=0,1 \quad (20)$$

만약 위의 식 (20)에서 이전 프레임에 음성구간이 있지 않을 때에는

$$\frac{p(S_i|H_n = H_1)}{p(S_i|H_n = H_0)} \underset{H_0}{\overset{H_1}{>}} \alpha' \frac{P(H_n = H_0|H_{n-1} = H_0)}{P(H_n = H_1|H_{n-1} = H_0)} = \beta', \quad (21)$$

이전 프레임에 음성구간이 있을 때에는

$$\frac{p(S_i|H_n = H_1)}{p(S_i|H_n = H_0)} \underset{H_0}{\overset{H_1}{>}} \alpha'' \frac{P(H_n = H_0|H_{n-1} = H_1)}{P(H_n = H_1|H_{n-1} = H_1)} = \beta'' \quad (22)$$

이러한 임계값 개수의 증가는 VAD의 성능이 향상될 수 있도록 추가적인 자유도를 제공한다. 인접한 프레임들간 음성 활동의 강력한 상호 연관성으로 인해 일반적으로 다음과 같은 식이 성립된다.

$$\frac{P(H_n = H_0|H_{n-1} = H_0)}{P(H_n = H_1|H_{n-1} = H_0)} > \frac{P(H_n = H_0|H_{n-1} = H_1)}{P(H_n = H_1|H_{n-1} = H_1)} \quad (23)$$

위의 식은 이전 프레임에서 음성 활동이 관측되었다면 현재 프레임에서 음성 활동이 관측될 확률이 크고 이전 프레임에서 음성 활동이 관측 되었을 때 현재 프레임에서 음성 활동이 관측되지 않을 확률이 작다. 또 그 반대의 경우도 가능하기 때문에 결론적으로  $\beta'$  이  $\beta''$  보다 크다는 것을 알 수 있다.

위의 결과들을 정리해 보면,

$$S_r(k, l) \underset{H_0}{\overset{H_1}{>}} \beta' \quad \text{if } H_{n-1} = H_0 \quad (24)$$

$$S_r(k, l) \underset{H_0}{\overset{H_1}{>}} \beta'' \quad \text{if } H_{n-1} = H_1 \quad (25)$$

위의 식을 간략히 표현을 하면 다음과 같이 나타낼 수 있다.

$$S_r(k, l) \underset{H_0}{\overset{H_1}{>}} \gamma \quad (26)$$

여기서

$$\gamma = \frac{P(H_{n-1} = H_0 | Y_{n-1})\beta}{(1 - P(H_{n-1} = H_0 | Y_{n-1}))\beta'} \quad (27)$$

위의 식처럼 이전 프레임의 음성신호의 유무에 따라 임계값에 변화를 주어 이를 가지고 음성 존재 확률을 유동적으로 추정하여 더 정확한 잡음 추정을 할 수가 있다.

그림 1은 MAP를 기반으로 한 임계값 테스트를 통한 음성 존재 확률을 보여 주고 있다. 기존의 MCRA방법보다 음성이 시작하는 부분은 더 빨리 음성임을 알아내고 음성이 끝나는 부분에서는 급격히 떨어지지 않아 음성임에도 불구하고 음성이 아니라고 판단해서 음성정보를 잃어버리는 것을 줄여 주는 것을 볼 수 있다.

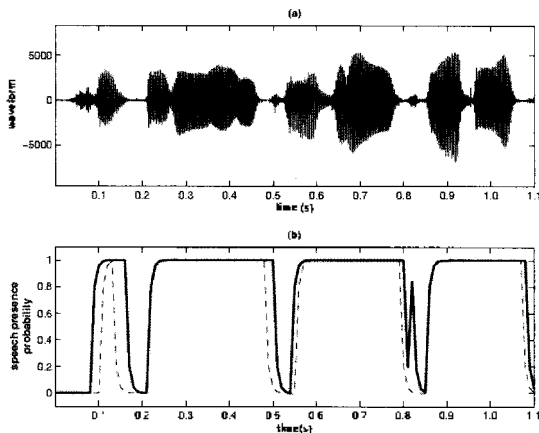


그림 1. F16 잡음 (SNR = 10 dB) 에서의 확률 비교 (a) 깨끗한 음성 파형 (b) 실시간 프레임에서의 음성 존재 확률: 기존의 MCRA의 확률 (점선), 제안된 알고리즘의 확률 (굵은선)

Fig. 1. Comparison of probability under the F16 noise (SNR = 10 dB) (a) Clean speech waveform (b) Speech presence probability in short-time frames: probability of conventional MCRA (dashed line), probability of proposed algorithm (bold line).

#### IV. 실험 결과

본 논문에서는 기존의 MCRA 알고리즘에서 각 서브 밴드에서 신호의 존재를 잡음이 섞인 신호의 국부에너지와 주어진 윈도우에서의 최소값 사이의 비로 정해 이 비율과 특정 임계값만을 가지고 비교를 하기 때문에 신뢰성이 떨어지는 문제를 조건 MAP를 적용함으로써 임계값을 지속적으로 갱신하여 음성 존재 확률값을 더 신뢰할 수 있도록 하여 음성 향상을 유도 하였다. 제안된 음성 향상 알고리즘의 음질 평가를 위해 널리 적용되고 있는 ITU-T

P.862 PESQ, 주관적 음질평가를 수행하였으며, 표 1과 표 2는 각각 추출된 PESQ수치와, 주관적 음질평가 비교를 보여주고 있다.

표 1의 PESQ 테스트를 위해 샘플은 남성, 여성화자 각각이 100개의 문장을 발음하도록 한 음성을 한 프레임의 크기가 10 ms에서 8 kHz로 샘플링 한 데이터에 네 가지 형태의 잡음이 부가 되었다. 잡음은 NOISEX-92 데이터 베이스의 white noise, babble noise, F16 noise를 사용하였으며 SNR으로 5, 10, 15 dB로 달리하여 테스트하였다. PESQ값은 이들 샘플에 대한 평균 수치로 나타냈고, 기존 MCRA에 의한 PESQ를 위해 가중치 파라미터  $\alpha_d=0.95$ ,  $\alpha_p=0.2$ ,  $\alpha_s=0.45$ 로 설정해 주었고, 임계값  $\beta'$ ,  $\beta''$  는 다양한 잡음환경에서 최적화된 실험치로 구하여  $\beta'=9$ ,  $\beta''=3.5$ 로 설정하였다.

표 1은 제안한 알고리즘과 기존의 MCRA를 비교하기 위해 PESQ 테스트를 실시한 결과 white noise와 F16 noise의 5 dB와 10 dB에서 두드러진 성능 향상을 보였다. 이는 제안한 알고리즘이 그림 1에서와 같이 낮은 SNR에서도 음성 존재 확률을 더욱 잘 추정함에 따라 개선된 음성 향상을 보이고 있음을 알 수 있다. babble 노이즈의 경우 잡음의 특성상 제안한 알고리즘의 음성 존재 확률이 기존과 비슷하지만 약간의 향상된 결과를 보여줌을 알 수 있다.

표 2의 주관적 음질평가는 남성, 여성 화자 각각이 10

표 1. 다양한 노이즈 환경에서 기존 MCRA와 향상된 MCRA의 PESQ 수치비교

Table 1. PESQ score of the MCRA and enhanced MCRA.

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	1.936	2.296	2.641
	Proposed	2.050	2.391	2.715
Babble noise	MCRA	2.379	2.697	2.972
	Proposed	2.397	2.724	3.004
F16 noise	MCRA	2.056	2.467	2.776
	Proposed	2.152	2.548	2.846

표 2. 다양한 노이즈 환경에서 기존 MCRA와 향상된 MCRA의 주관적 음질평가 비교

Table 2. The subjective evaluation of speech quality of the MCRA and enhanced MCRA.

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	2.17	2.56	3.03
	Proposed	2.23	2.65	3.07
Babble noise	MCRA	2.93	3.32	3.69
	Proposed	2.95	3.42	3.76
F16 noise	MCRA	2.23	2.77	3.20
	Proposed	2.33	2.90	3.46

개의 문장을 발음하도록 한 음성에 white, babble, F16 noise가 SNR이 5, 10, 15 dB로 부과된 잡음 신호를 20명의 청취자를 대상으로 평가하였다. 표 2는 세 가지 잡음 환경과 각 SNR에 대하여 모두 향상된 결과를 보여주고 있다. 결론적으로, 제안된 조건부 MAP를 기반으로 한 MCRA 알고리즘이 다양한 잡음 환경에서 우수함을 알 수 있었다.

### V. 결론

본 논문에서는 기존의 MCRA 알고리즘에 조건 MAP를 사용하여 음성 존재 확률을 더욱 정확하게 추정하는 새로운 알고리즘을 제안하였다. 기존의 방법은 효율적인 지역 최소 추적 기술을 사용하여 계산의 복잡성은 줄였지만 갑작스러운 잡음이 존재할 때 딜레이가 생기게 된다. 또한 각 서브밴드에서 신호의 존재를 잡음이 섞인 신호의 국부에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한다. 이 비율과 특정 한계 값만 가지고 비교를 하기 때문에 신뢰성이 떨어지는 단점이 있었다.

본 논문에서 이러한 특정 임계값만을 사용하여 각 서브밴드에서 음성신호의 유무를 추정할 때 이전 서브밴드의 신호가 음성을 포함하고 있는지 아닌지에 대한 조건을 추가하여 더욱 신뢰성 있는 비교를 통해 신호 존재 확률의 성능을 향상시켰으며 그 결과 실험에 사용된 모든 잡음과 신호 대 잡음 비 환경에서 기존의 MCRA 보다 음성 향상에서 우수한 성능을 보였다.

### 감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업 [2008-F-045-01] 과 지식경제부 및 정보통신연구진흥원의 대한 IT 연구센터 지원사업의 연구 결과로 수행되었음 (IITA-2008-C1090-0804-0007).

### 참고 문헌

1. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-32(6), 1109-1121, Dec. 1984.
2. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-32(2), 443-445, Apr. 1985.

3. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-27(2), 113-120, Apr. 1979.
4. N. S. Kim and J. H. Chang, "Spectral enhancement based on global soft decision," IEEE Signal Processing Letters, 7(5), 108-110, May 2000.
5. R. Martin, "Spectral subtraction based on minimum statistics," Proceeding of 7th EUSIPCO'94, Edinburgh, U.K., 1182-1185, Sept. 1994.
6. I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Processing Letters, 9(1), 12-15, Jan. 2002.
7. J. W. Shin, H. J. Kwon, S. H. Jin and N. S. Kim, "Voice activity detection based on conditional MAP criterion," IEEE Signal Processing Letters, 15, 257-260, Feb. 2008.
8. I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," Signal Processing, 81, 2403-2418, Nov. 2001.
9. G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," Proceeding of 4th EUROSpeech'95, Madrid, Spain, 1513-1516, Sept. 1995.
10. J. Meyer, K. U. Simmer and K. D. Kammeyer, "Comparison of one- and two-channel noise-estimation techniques," Proceeding of 5th IWAENC'97, London, U.K., pp. 137-145, Sept. 1997.
11. J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection," IEEE Signal Processing Letters, 6(1), 1-3, Jan. 1999.

### 저자 약력

#### •김종모 (Jong-Mo Kum)



2008년 2월: 인하대학교 전자공학과 학사  
2008년 3월~현재: 인하대학교 전자공학부 석사과정

#### •박윤식 (Yun-Sik Park)



2006년 2월: 인하대학교 전자공학과 학사  
2008년 2월: 인하대학교 전자공학과 석사  
2008년 3월~현재: 인하대학교 전자공학부 박사과정

#### •장준혁 (Joon-Hyuk Chang)



1998년 2월: 경북대학교 전자공학과 학사  
2000년 2월: 서울대학교 전기공학부 석사  
2004년 2월: 서울대학교 전기컴퓨터공학부 박사  
2000년 3월~2005년 4월: ㈜넷디스 연구소장  
2004년 5월~2005년 4월: 캘리포니아 주립대학, 산타바바라 (UCSB) 박사후연구원  
2005년 5월~2005년 8월: 한국과학기술연구원 (KIST) 진영연구원  
2005년 9월~현재: 인하대학교 전자공학부 조교수