논문 2009-46SP-4-17

2차 조건 사후 최대 확률 기반 최소값 제어 재귀평균기법을 이용한 음성향상

(Speech Enhancement based on Minima Controlled Recursive Averaging Technique Incorporating Second-order Conditional Maximum a *posteriori* Criterion)

금 종 모*, 장 준 혁**

(Jong-Mo Kum and Joon-Hyuk Chang)

요 약

본 논문에서는 기존의 Minima Controlled Recursive Averaging (MCRA)에 2차 조건 사후 최대 확률기법을 적용한 음성 향상 기법을 제안한다. 기존의 MCRA 방법은 현재 프레임의 음성 신호 존재 확률로 잡음 추정을 조정하기 때문에 음성 활동의 프레임간의 상호 연관성을 배제 하였다. 본 논문에서 제안한 알고리즘은 직전 2 프레임에서의 음성의 존재와 부재에 대한 조건을 부여해 주어 현재 프레임의 음성 신호 존재 확률을 수정하는 음성향상 기법을 적용한다. 제안된 2차 조건 사후 최대 확률기법을 적용한 MCRA 방법이 기존의 MCRA 방법보다 향상된 음성향상 결과를 나타내었다.

Abstract

In this paper, we propose a novel approach to improve the performance of minima controlled recursive averaging (MCRA) which is based on the second-order conditional maximum a posteriori (CMAP). From an investigation of the MCRA scheme, it is discovered that the MCRA method cannot take full consideration of the inter-frame correlation of voice activity since the noise power estimate is adjusted by the speech presence probability depending on an observation of the current frame. To avoid this phenomenon, the proposed MCRA approach incorporates the second-order CMAP criterion in which the noise power estimate is obtained using the speech presence probability conditioned on both the current observation and the speech activity decisions in the previous two frames. Experimental results show that the proposed MCRA technique based on second-order conditional MAP yields better results compared to the conventional MCRA method.

Keywords: Speech Enhancement, Minima Controlled Recursive Averaging (MCRA), Second-order Conditional Maximum *a posteriori* (Second-order CMAP)

Ⅰ. 서 론

이동환경에서의 음성 통신의 중요성이 점차 증가하면서 단일 마이크로폰에서의 음성향상 기술에 대한 연구가 주목받고 있다. 실제적인 음성향상 시스템에서 잡음을 정확하게 추정하는 것은 핵심 요소 이며, 다양한 환경 특히 비정상 잡음신호나 간섭 신호 또는 신호 대 잡음비가 낮은 신호를 처리할 수 있어야 한다[1~9]. 최근에 주목받는 잡음 추정 방법으로 최소값 제어 재귀평균(Minima Controlled Recursive Averaging, MCRA)기법은 서브밴드에서 신호 존재 확률로 조절하는 스무딩 매개변수를 이용하여 파워 스펙트럼에 평균을 취하는 방

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

※ 본 연구는 지식경제부 및 정보통신연구진흥원의 IT 핵심기술개발사업 [2008-F-045-01]과 본 연구는 지식경제부와 한국산업기술재단의 전략기술인력양성사업으로 수행된 연구결과임

접수일자: 2008년10월22일, 수정완료일: 2009년6월4일

법이다[10~13]. 각 서브 밴드에서 신호의 존재는 잡음이 섞인 신호의 국부 에너지와 주어진 윈도우에서의 최소값 사이의 비와 특정 임계값과의 비교를 통하여 비율이 작으면 음성신호가 없는 것으로 결정하며, 음성 신호가 있는 부분과 없는 부분 사이에 발생하는 변동을 줄이기 위해 시간축 으로도 평균을 취한다. 하지만 이러한 MCRA 알고리즘에도 몇 가지의 문제점들이 있다. 특히 효율적인 지역 최소 추적 기술을 사용하여 계산의 복잡성은 줄였지만 갑작스러운 잡음 존재시 딜레이가 생기고, 각 서브샌드에서 신호의 존재를 잡음이 섞인 신호의 국부에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한 값과 특정 임계값만을 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 반대 경우도 성립된다. 최근에[9] 제안된 음성 활동의 상호 연관성을 고려한 새로운 음성 검출기 (voice activity detection, VAD)에서는 조건 사후 최대 확률 (conditional maximum *a posteriori*, CMAP)기법을 적용하여 우수한 성능을 보고하였다.

본 논문에서는 특정 임계값만을 사용하여 각 서브밴드에서 음성신호의 유무를 추정하는 기존의 MCRA 방법에 이전 두 프레임의 음성신호 유무에 대한 조건을 추가한 2차 조건 사후 최대 확률 (second-order CAMP)을 적용하여 음성 존재 확률의 성능을 향상시켜 우수한 잡음 추정 방법을 도출하여 향상된 MCRA를 제시한다. 제안된 음성 향상 기법은 ITU-T P.862 perceptual evaluation of speech quality (PESQ)[15]와 주관적 음질 평가를 통해 평가 했고 기존의 MCRA보다 향상된 결과를 나타내었다.

II. MCRA (minima controlled recursive averaging)

$x(n)$과 $d(n)$을 각각 음성신호와, 상관성이 없는 가산 잡음 신호라고 한다. 여기서 n은 이산시간을 나타낸 다면, 관측되는 신호 $y(n)$은 $y(n)=x(n)+d(n)$으로 주어지고 부분적으로 중복으로 나누어 윈도우를 취한 다음 단구간 퓨리에 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 l번째 프레임에서의 k번째 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 퓨리에 변환 계수를 나타낸다. 여기서, $\lambda_d(k,l)=E[|D(k,l)|^2]$이 k번째 서브밴드에서 잡음 신호의 분산이라고 하면, 추정하기 위해 음성 신호 부재 구간에서 관측된 신호에 시간의 반복 스무딩을 적용하면 다음과 같이 나타낼 수 있다.

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

여기서 $\alpha_d(0<\alpha_d<1)$는 스무딩 매개변수이다. H'_0, H'_1은 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 위험하다고 고려된다. 그러므로 서로 다른 결정 법칙이 사용되고, 일반적으로 H'_1보다는 H_1에 더 높은 신뢰를 두고 있으며, $P(H_1|Y)\geq P(H'_1|Y)$인 것이다[10].

$p(k,l)=P(H'_1(k,l)|Y(k,l))$가 음성 존재의 조건 확률을 나타낸다고 하면, 식 (3)은 다음과 같다.

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

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

$$\hat{\alpha}_d(k,l)=\alpha_d+(1-\alpha_d)P(H'_1(k,l)|Y(k,l)). \tag{5}$$

$P(H'_1(k,l)|Y(k,l))$에 대한 추정은 다음과 같은 식을 사용

한다[6].

$$P(H_1'(k,l)|Y(k,l)) = \begin{cases} \alpha_p P(H_1'(k,l-1)|Y(k,l-1)) \\ +(1-\alpha_p), \quad \text{if } S_r(k,l) > \delta \\ \alpha_p P(H_1'(k,l-1)|Y(k,l-1)), \\ \qquad otherwise \end{cases} \quad (6)$$

여기서 $\alpha_p(0<\alpha_p<1)$는 스무딩 매개변수이고, δ는 음성신호 존재의 임계값이다. 또한 $S_r(k,l) = S(k,l)/S_{\min}(k,l)$은 잡음이 섞인 신호의 국부에너지, $S(k,l)$와 주어진 윈도우에서의 최소값, $S_{\min}(k,l)$의 비를 나타낸다. 여기서 $S_r(k,l)$은 아래의 Bayes의 최고 비용 결정 방법을 기반으로 하고 있다.

$$\frac{P(S_r(k,l)|H_1(k,l))}{P(S_r(k,l)|H_0(k,l))} \underset{H_0'}{\overset{H_1'}{\gtrless}} \alpha \frac{C_{10}P(H(k,l)=H_0)}{C_{01}P(H(k,l)=H_1)} \quad (7)$$

여기서 $P(H(k,l)=H_0) = (1-P(H(k,l)=H_1))$는 음성 부재의 *a priori* 확률이고, C_{ij}는 H_j'일 때 H_i'인 결정 비용이다. 식 (7)은 간단한 함수이기 때문에 결정식 (7)은 식 (6)처럼 표현 할 수 있다.

$$S_r(k,l) \underset{H_0'}{\overset{H_1'}{\gtrless}} \delta \quad (8)$$

여기서 $\delta = \alpha \frac{C_{10}P(H(k,l)=H_0)}{C_{01}P(H(k,l)=H_1)}$ 이다. $S(k,l)$을 얻기 위해 다음식과 같이 재귀평균을 이용하였다.

$$S(k,l) = \zeta_s S(k,l-1) + (1-\zeta_s)|Y(k,l)|^2 \quad (9)$$

여기서 $\zeta_s(0<\zeta_s<1)$은 스무딩 파라미터이다. 추가적으로, 현재 프레임의 최소값 $S_{\min}(k,l)$를 얻기 위해서 이전 프레임의 최소값과 국부에너지를 비교하였다[10].

III. 제안된 2차 조건 MAP (maximum *a posteriori*)를 기반으로 한 향상된 MCRA

지금까지 우리는 MCRA 방법에서 잡음이 섞인 신호에서 재귀평균을 이용한 국부에너지와 이전 프레임의 최소값의 비로 조절되는 중요한 파라미터 $P(H_1'(k,l)|Y(k,l))$에 대해 알아보았다. 하지만 $P(H_1'(k,l)|Y(k,l))$는 음성 활동에서 인접한 프레임들의 상호 연관성을 고려하지 않았다. 이러한 상호 연관성은 HMM을 이용한 행오버를 사용함으로서 통계모델을 기반으로 한 VAD의 에러를 효과적으로 줄일 수 있다[14]. 음성 활동에서 프레임들간의 강력한 상호 연관성에 기반 하여 다음과 같이 표현할 수 있다.

$$P(H(k,l)=H_1|H(k,l-1)=H_1, H(k,l-2)=H_1) > P(H(k,l)=H_1) \quad (10)$$

첫 번째로 고려할 사항은 현재 프레임의 관찰 결과와 이전 $(l-1)$ 프레임과 그 이전 $(l-2)$ 프레임에서의 아래와 같은 음성 존재 결정 결과이다.

$$\frac{p(H(k,l)=H_1|S_r,H(k,l-1)=H_i,H(k,l-2)=H_j)}{p(H(k,l)=H_0|S_r,H(k,l-1)=H_i,H(k,l-2)=H_j)} \underset{H_0'}{\overset{H_1'}{\gtrless}} \alpha$$

i=0,1 j=0,1. (11)

Bayes의 룰을 사용하여 식 (7)과 같은 형태로 위의 식을 다음과 같이 바꾸어 줄수 있다.

$$\frac{p(S_r|H(k,l)=H_1,H(k,l-1)=H_i,H(k,l-2)=H_j)}{p(S_r|H(k,l)=H_0,H(k,l-1)=H_i,H(k,l-2)=H_j)} \underset{H_0'}{\overset{H_1'}{\gtrless}} \alpha_{ij}'$$

i=0,1 j=0,1 (12)

여기서

$$\alpha_{ij}' = \alpha \frac{P(H(k,l)=H_0|H(k,l-1)=H_i,H(k,l-2)=H_j)}{P(H(k,l)=H_1|H(k,l-1)=H_i,H(k,l-2)=H_j)}$$

이다. 비록 현재 프레임의 음성 활동이 이전 프레임들에 의존할 지라도, 현재 프레임의 음성 활동은 현재 프레임에서 관찰된 잡음 섞인 음성 신호의 DFT계수의 분포에 지배적인 영향을 받는다. 그러므로 식 (12)은 다음과 같이 간단히 표현할 수 있다.

$$\frac{p(S_r|H(k,l)=H_1)}{p(S_r|H(k,l)=H_0)} \underset{H_0'}{\overset{H_1'}{\gtrless}} \alpha_{ij}' \quad \text{i=0,1 j=0,1} \quad (13)$$

여기서

$$\alpha_{00}' \left(= \frac{P(H(k,l)=H_0|H(k,l-1)=H_0,H(k,l-2)=H_0)}{P(H(k,l)=H_1|H(k,l-1)=H_0,H(k,l-2)=H_0)}\right)$$

은 이전 프레임에 음성이 존재하지 않고 그 이전 프레임

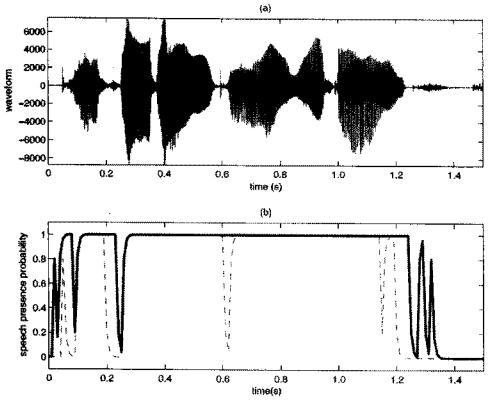

그림 1. F16 잡음 (SNR = 10 dB) 에서의 확 률 비교 (a) 깨끗한 음성 파형 (b) 실시간 프레임에서의 음성 존재 확률: 기존의 MCRA의 확률 (점선), 제안된 알고리즘의 확률 (굵은선).

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).

에도 음성이 존재하지 않을 때 이며,

$$\alpha'_{01}\left(=\frac{P(H(k,l)=H_0|H(k,l-1)=H_0,H(k,l-2)=H_1)}{P(H(k,l)=H_1|H(k,l-1)=H_0,H(k,l-2)=H1)}\right)$$

은 이전 프레임에 음성이 존재하지 않고 그 이전 프레임에 음성이 존재할 때 이다. 또한

$$\alpha'_{10}\left(=\frac{P(H(k,l)=H_0|H(k,l-1)=H_1,H(k,l-2)=H_0)}{P(H(k,l)=H_1|H(k,l-1)=H_1,H(k,l-2)=H_0)}\right)$$

은 이전 프레임에는 음성이 존재하고 그 이전 프레임에는 음성이 존재하지 않을 때 이며,

$$\alpha'_{11}\left(=\frac{P(H(k,l)=H_0|H(k,l-1)=H_1,H(k,l-2)=H_1)}{P(H(k,l)=H_1|H(k,l-1)=H_1,H(k,l-2)=H_1)}\right)$$

은 이전 프레임과 그 이전 프레임에 모두 음성이 존재 할 때 이다. 이러한 임계값 개수의 증가는 VAD의 성능의 향상이 될 수 있도록 추가적인 자유도를 제공한다. 위의 식처럼 이전 프레임과 그 이전 프레임의 음성 신호의 유무에 따라 임계값에 변화를 주어 이를 가지고 음성 존재 확률을 유동적으로 추정하여 더 정확한 잡음 추정을 할 수가 있다. 위의 식들을 간단히 하면 다음과 같이 나타낼 수 있다.

$$S_r(k,l) \underset{H_0}{\overset{H_1}{\gtrless}} \alpha'_{ij} \qquad \text{i=0,1 j=0,1} \qquad (14)$$

그림 1은 2차 CMAP를 기반 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.

실제로 제안된 기법은 minimum mean square error (MMSE)를 기반으로 한 음성 향상 알고리즘에 주로 이용된다[1, 8].

$$\hat{X}(k,l) = G(\xi(k,l),\gamma(k,l))\,Y(k,l) \qquad (15)$$

여기서 $\hat{X}(k,l)$은 추정된 깨끗한 음성 신호 이고, $G(\bullet)$은 잡은 제거 이득이다. 또한 $\gamma(k,l)$과 $\xi(k,l)$은 각각 a posteriori SNR, a priori SNR로 정의 된다.

$$\gamma(k,l) \equiv \frac{|Y(k,l)|}{\lambda_d(k,l)}, \qquad (16)$$

$$\xi(k,l) \equiv \frac{\lambda_x(k,l)}{\lambda_d(k,l)}. \qquad (17)$$

여기서 MMSE 잡은 제거 이득은 다음과 같이 주어진다[1].

$$G(\hat{\xi}(k,l),\hat{\gamma}(k,l)) = \frac{\sqrt{\pi\nu(k,l)}}{2\hat{\gamma}(k,l)} epx\left(-\frac{\nu(k,l)}{2}\right)$$
$$\bullet\ [(1+\nu(k,l))I_0\left(\frac{\nu(k,l)}{2}\right)$$
$$+\nu(k,l)I_1\left(\frac{\nu(k,l)}{2}\right)] \qquad (18)$$

여기서 I_0와 I_1은 각각 영차와 일차 Bessel 함수이다. 또한 $\nu(k,l)$은 (4), (16), (17)을 기반으로 다음과 같이 정의된다.

$$\nu(k,l) = \frac{\hat{\xi}(k,l)}{1+\hat{\xi}(k,l)}\hat{\gamma}(k,l) \qquad (19)$$

여기서 $\hat{\gamma}(k,l)$과 $\hat{\xi}(k,l)$은 (16), (17)을 사용한 고유 추정치 이다.

IV. 실험 결과

제안된 음성 향상 알고리즘의 음질 평가를 위해 널리 적

표 1. 다양한 노이즈 환경에서 기존 MCRA와 향상된 MCRA의 Relative Estimation Error 비교.
Table 1. Relative Estimation Error from the MCRA and enhanced MCRA.

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	0.379	0.401	0.607
	Proposed	0.369	0.386	0.425
Babble noise	MCRA	0.892	0.932	1.593
	Proposed	0.714	0.717	0.756
F16 noise	MCRA	0.457	0.688	1.444
	Proposed	0.348	0.357	0.392

용되고 있는 relative estimation error, ITU-T P.862 PESQ, 주관적 음질평가를 수행하여 음성향상 비교를 하였다.

첫 번째로 잡은 추정 성능을 평가 할 수 있는 정규화된 relative estimation error ϵ_n 는 다음과 같이 정의 된다[10].

$$\epsilon_n = \frac{1}{N}\sum_{t=0}^{N-1}\frac{\sum_k\left[\hat{\lambda}_d(k,l)-\lambda_d(k,l)\right]^2}{\sum_k\lambda_d^2(k,l)} \qquad (20)$$

여기서 $\lambda_d(k,l)$은 잡음 신호에 의해 직접적으로 얻어지는 실제 잡음 추정치이고[10], $\hat{\lambda}_d(k,l)$은 N번째 프레임에서의 제안된 방법을 사용한 향상된 MCRA를 통한 잡음 추정치 이다. 표 1은 다양한 잡음 환경에서의 잡음 추정을 위한 relative estimation error의 결과를 보여주고 있다. 이 결과 제안된 방법이 이전의 MCRA방법보다 보다 나은 성능을 보여주고 있다[10].

표 2의 ITU-T P.862 perceptual evaluation of speech quality (PESQ)[15] 테스트를 위해 샘플은 남성, 여성화자 각각이 100개의 문장을 발음하도록 한 음성을 한 프레임의 크기가 10 ms에서 8 kHz로 샘플링 한 데이터에 세 가지 형태의 잡음이 부가 되었다. 잡음은 NOISEX-92 데이터베이스의 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 α_d=0.95, α_p=0.2, α_s=0.45로 설정해 주었고, 임계값 α'_{00}, α'_{01}, α'_{10}, α'_{11}은 다양한 잡음환경에서 최적화된 실험치로 구하여 α'_{00}=5, α'_{01}=4, α'_{10}=1.5, α'_{11}=1.1로 설정하였다.

표 2. 다양한 노이즈 환경에서 기존 MCRA와 향상된 MCRA의 PESQ 수치비교.
Table 2. 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	2.066	2.399	2.723
Babble noise	MCRA	2.315	2.645	2.927
	Proposed	2.316	2.649	2.956
F16 noise	MCRA	2.029	2.442	2.757
	Proposed	2.182	2.541	2.841

표 3. 다양한 노이즈 환경에서 기존 MCRA와 향상된 MCRA의 주관적 음질평가 비교
Table 3. The subjective evaluation of speech quality of the MCRA and enhanced MCRA.

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	1.97	2.69	2.94
	Proposed	2.16	2.81	3.13
Babble noise	MCRA	2.96	3.40	3.66
	Proposed	3.05	3.42	3.70
F16 noise	MCRA	2.09	2.74	3.28
	Proposed	2.30	2.87	3.41

표 2는 제안한 알고리즘과 기존의 MCRA 를 비교하기 위해 PESQ 테스트를 실시한 결과 white noise와 F16 noise 의 5dB 와 10 dB 에서 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 3의 주관적 음질평가는 남성, 여성 화자 각각이 10개의 문장을 발음하도록 한 음성에 white, babble, F16 noise가 SNR이 5, 10, 15 dB로 부과된 잡음 신호를 10명의 청취자를 대상으로 평가하였다. 각각의 청취자들은 각 문장을 듣고 5 (Excellent), 4 (Good), 3 (Fair), 2 (Poor), 1 (Bad)의 점수를 주었다. 표 2는 세

가지 잡음 환경과 각 SNR에 대하여 10명의 평가자들의 평균을 낸 mean opinion score 이고, 결과적으로 모두 향상된 것을 보여주고 있다. 제안된 2차 조건 MAP를 기반으로 한 MCRA 알고리즘이 다양한 잡음 환경에서 이전의 MCRA방법 보다 우수함을 알 수 있다.

V. 결 론

본 논문에서는 기존의 MCRA 알고리즘에 2차 조건 MAP를 사용하여 음성 존재 확률을 더욱 정확하게 추정하는 새로운 알고리즘을 제안하였다. 잡음의 추정은 현재 프레임의 관찰 결과와 이전 두 프레임의 음성 신호 유무의 결정 결과로서 조정 할 수 있는 음성 존재 확률을 파라미터를 사용하여 재귀 평균을 통해 구하였다. 제안된 방법은 기존의 MCRA 방법과 비교를 해본 결과 음성 향상 시스템을 위한 잡음 추정이 보다 향상되는 것을 보였다.

참 고 문 헌

[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*, vol. ASSP-32, no. 6, pp. 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*, vol. ASSP-32, no. 2, pp. 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*, vol. ASSP-27, no. 2, pp. 113-120, Apr. 1979.

[4] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. EUSIPCO*, Edinburgh, U.K., pp. 1182-1185, Sept. 1994.

[5] G. Doblinger, "Computationally efcient speech enhancement by spectral minima tracking in subbands," in *Proc. EUROSPEECH*, Madrid, Spain, pp. 1513-1516, Sept. 1995.

[6] J. Meyer, K. U. Simmer and K. D. Kammeyer, "Comparison of one-and two-channel noise-estimation techniques," in *Proc. IWAENC*, London, U.K., pp. 137-145, Sept. 1997.

[7] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, pp. 2403-2418, Nov. 2001.

[8] N. S. Kim and J.-H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, no. 5, pp. 108-110, May 2000.

[9] 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*, vol. 15, pp. 257-260, Feb. 2008.

[10] I. Cohen and 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.

[11] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466-475, Sept. 2003.

[12] V. Stouten, H. V. hamme, P. Wambacq, "Application of minimum statistics and minima controlled recursive averaging methods to estimate a cepstral noise model for robust ASR," in *Proc. ICASSP*, Toulouse, France, pp. 765-768, May. 2006.

[13] N. Fan, J. Rosca, R. Balan, "Speech noise estimation using enhanced minima controlled recursive averaging," in *Proc. ICASSP*, Honolulu, Hawaii, U.S.A., pp. 581-584, Apr. 2007.

[14] J. Sohn, N. S. Kim andW. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol. 6, no 1, pp. 1-3, Jan. 1999.

[15] ITU-T P.862, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, 2001.

저 자 소 개

금 종 모(학생회원)
2008년 인하대학교 전자전기 공학부 학사.
2008년~현재 인하대학교 전자공학과 석사과정.
<주관심분야 : 디지털신호처리>

장 준 혁(정회원)
1998년 경북대하교 전자공학과 학사.
2000년 서울대학교 전기공학부 석사.
2004년 서울대학교 전기컴퓨터 공학부 박사.
2000년~2005년 (주)넷더스 연구소장
2004년~2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
2005년 한국과학기술연구원(KIST) 연구원
2005년~현재 인하대학교 전자공학부 조교수
<주관심분야 : 음성 신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>