

잡음 환경 분류 알고리즘을 이용한 IMCRA 기반의 음성 향상 기법

Speech Enhancement Based on IMCRA Incorporating noise classification algorithm

송지현* · 박규석* · 안홍섭* · 이상민†
(Ji-Hyun Song · Gyu-Seok Park · Hong-Sub An · Sang-Min Lee)

Abstract - In this paper, we propose a novel method to improve the performance of the improved minima controlled recursive averaging (IMCRA) in non-stationary noisy environment. The conventional IMCRA algorithm efficiently estimate the noise power by averaging past spectral power values based on a smoothing parameter that is adjusted by the signal presence probability in frequency subbands. Since the minimum of smoothing parameter is defined as 0.85, 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. For this reason, we proposed the modified IMCRA, which adaptively estimate and update the noise power according to the noise type classified by the Gaussian mixture model (GMM). The performances of the proposed method are evaluated by perceptual evaluation of speech quality (PESQ) and composite measure under various environments and better results compared with the conventional method are obtained.

Key Words : Improved minima controlled recursive averaging(IMCRA), Speech enhancement, GMM

1. 서론

인간 생활의 편의를 위해 다양한 기술들이 개발되고 있고, 그 중 음성 검색, 음성 인식기와 같이 음성 신호를 이용하는 기술이 많은 관심을 받고 있다. 이러한 음성 시스템에서 배경 잡음의 존재는 시스템의 성능저하를 발생시키는 가장 큰 요인으로 이를 제거하기 위한 음성 향상 기술 연구가 진행되었다 [1]-[3].

음성 향상 기술은 크게 잡음 전력을 추정하는 부분과 이를 이용하여 각 주파수에 대한 이득을 추정하는 부분으로 나눌 수 있다. 신호의 잡음 전력이 원 잡음 전력 보다 크게 추정되면 잡음 신호 뿐만 아니라 음성 신호 또한 제거되어 왜곡이 발생하고, 이와 반대로 잡음 전력이 원 잡음 전력보다 작게 추정되면 잡음 신호가 덜 제거되어 음성의 명료도가 감소하게 되므로 정확한 잡음 전력 추정은 음성 향상 시스템에서 중요한 요소이다 [4],[5]. 이러한 잡음 전력을 추정하는 기술로 코헨에 의해서 제안된 minima controlled recursive averaging (MCRA) 방법과 improved minima controlled recursive averaging (IMCRA) 방법이 우수한 성능을 보인다고 알려져 있다 [6],[7]. MCRA 방법은 음성 신호와 잡음 신호가 통계적으로 독립이라는 가정 하에 잡음이 부가된 신호의 전력값이 자주 잡음 신호의 전력 값까지 감소한다는 관찰을 기반으로 한다. 적절한 윈도우 내에서 최

소값을 구하고 구해진 최소값과 입력된 신호의 비를 정해진 문턱 값과의 비교를 통해서 음성의 존재 유무를 판단하고, 이 결과를 이용하여 잡음 전력의 업데이트에 사용되는 스무딩 파라미터의 값이 결정되어 잡음 전력이 추정된다. IMCRA 방법은 MCRA 방식을 변형한 알고리즘으로 크게 두 가지 측면에서 차이점을 볼 수 있다 [7]. 첫 번째, MCRA 방법의 음성 유무를 판단하는 알고리즘을 이용하여 음성 특성이 강한 음성 구간의 주파수 스펙트럼을 이전 프레임의 잡음 주파수 스펙트럼의 스무딩 된 값으로 대체하고, 다시 한 번 MCRA 방식을 실행하여 잡음 전력을 추정한다. 즉, 강한 음성 구간의 서브 밴드에서 잘못된 잡음 전력 업데이트가 발생하는 것을 방지하여 보다 정확한 잡음 전력을 추정하게 해준다. 두 번째는 MCRA에서 음성 유무를 판단하는 알고리즘에서 사용되는 특징벡터의 추가 이다. 기존의 윈도우내의 최소전력과 현재 입력된 전력의 비율 외에 윈도우내의 최소전력과 시간축-주파수 축으로 스무딩된 전력의 비율을 추가로 이용하였다. 이러한 두 가지 측면에 의해서 기존의 MCRA 방식에 비해서 건설한 잡음 전력을 추정한다. 하지만 잡음 전력 업데이트에 사용되는 스무딩 파라미터의 경우 $\alpha_d = 0.85 \sim 1$ 의 범위를 갖기 때문에 비정상 잡음과 같이 주파수 특성이 상대적으로 빠르게 변하는 구간에서는 성능저하를 유발한다.

본 논문에서는 비정상적인 잡음 환경에서 IMCRA의 보다 정확한 잡음 전력 추정을 위해서 Gaussian mixture model (GMM) 기반의 잡음 환경 알고리즘을 이용한 향상된 음성 향상 기법을 제안한다. 구체적으로, IMCRA 알고리즘에 의해서 잡음 구간이라고 판단될 경우 잡음 환경 알고리즘을 통해서 잡음 환경이 분류 되고, 그 결과에 따라서 잡음 전력 추정에 사용되는 스무딩 파라미터를 비정상적 잡음 환경에

* Dept. of Electrical Engineering, Inha University

† Corresponding Author : Dept. of Electrical Engineering, Inha University

E-mail : sanglee@inha.ac.kr

Received : August 23, 2012; Accepted : November 28, 2012

서는 현재 프레임의 잡음 전력에 더 큰 가중치가 적용되게 하고, 정상적 잡음 환경에서는 이전 프레임의 잡음 전력에 더 큰 가중치가 적용되도록 하였다. 즉, 잡음 신호의 종류에 따라서 적응적으로 잡음 전력 추정에 사용되는 스무딩 파라미터의 값이 변하도록 하여 보다 견실한 잡음 전력을 추정하는 알고리즘을 제안하였다. 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.

본 논문의 2장에서는 IMCRA, 3장에서는 잡음 환경 분류 알고리즘 기반의 제안된 알고리즘에 대해서 살펴본다. 그리고 4장에서 실험 결과를 종합적으로 검토하고 5장에서 결론을 맺는다.

2. IMCRA (Improved minima controlled recursive averaging) 개요

음성 신호 $x(n)$ 에 가산 잡음 신호 $d(n)$ 가 인가된 오염된 음성 신호 $y(n)$ 는 음성 존재 $H_1(k,l)$ 와 음성 부재 $H_0(k,l)$ 에 대해서 다음과 같이 나타낼 수 있다.

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

여기서 k 는 주파수 인덱스, l 은 시간축 인덱스를 나타낸다. 위의 가설을 기반으로 잡음 전력 추정 방식은 VAD를 이용하여 잡음만 존재하는 상황에서 잡음 전력을 추정하고 업데이트 하는 hard decision 방식과 음성 존재 확률을 스무딩 파라미터로 이용하여 잡음 전력을 추정하는 soft decision 방식으로 나눌 수 있다. IMCRA 알고리즘은 후자인 음성 존재 확률을 스무딩 파라미터로 이용하여 잡음 전력을 추정하고, 다음과 같이 잡음 전력을 추정한다. 먼저, VAD를 이용한 잡음 추정 식을 나타내면 다음과 같다.

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

여기서 α_d 는 시간축 스무딩 매개 변수를 나타내고, $\bar{\lambda}_d(k,l)$ 는 잡음 전력을 나타낸다. IMCRA에서는 식(2)에 음성 존재 확률 ($p'(k,l) = p(H_1(k,l)|\gamma(k,l))$)을 적용하여 다음과 같이 잡음 전력을 추정한다.

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

여기서 $p'(k,l) = p(H_1(k,l)|\gamma(k,l))$ 는 a posteriori SNR의 조건부 음성 존재 확률을 나타내고, $\bar{\alpha}_d(k,l)$ 는 음성 존재 확률에 의해서 변화되는 스무딩 매개 변수로 다음과 같이 정의 된다.

$$\bar{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d) p'(k,l) \quad (4)$$

여기서 α_d 는 고정된 스무딩 변수를 나타낸다. 현재 프레임의 음성 존재 확률이 작을 경우 이전 프레임에서 추정된 잡음 전력에 가중치 ($\bar{\alpha}_d(k,l) \approx 0.85$)를 주고, 반대의 경우 현재 프레임에서 추정된 잡음 전력에 가중치 ($1-\bar{\alpha}_d(k,l) \approx 0.15$)를 준다. 만약 주파수 스펙트럼이 시간에 따라서 빠르게 변화하는 비정상적인 잡음 상황에서는 잡음 전력 추정의 느린 업데이트에 의해서 음질 저하가 발생하게 될 여지가 존재한다.

a posteriori SNR의 조건부 음성 존재 확률은 다음과 같이 구해진다.

$$\begin{aligned} p'(k,l) &= p(H_1(k,l)|\gamma(k,l)) \\ &= \left\{ 1 + \frac{q(k,l)}{1-q(k,l)} (1 + \xi(k,l)) \exp\left(-\frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}\right) \right\}^{-1} \end{aligned} \quad (5)$$

여기서 $q(k,l)$ 는 사전 음성 부재 확률을 나타내고, $\gamma(k,l) = |Y(k,l)|^2/\lambda_d(k,l)$, $\xi(k,l) = \lambda_s(k,l)/\lambda_d(k,l)$ 은 각각 a posteriori SNR과 a priori SNR을 나타낸다.

사전 음성 부재 확률은 다음과 같은 과정을 통해서 구해진다. 먼저 입력된 신호 $Y(k,l)$ 를 주파수-시간 축으로 스무딩을 한다.

$$\begin{aligned} S_f(k,l) &= \sum_{i=-w}^w b(i) |Y(k-1,l)|^2 \\ S(k,l) &= \alpha_s S(k,l-1) + (1-\alpha_s) S_f(k,l) \end{aligned} \quad (6)$$

여기서 $b(i)$ 는 $2w-1$ 길이의 해밍윈도우를 나타내고, α_s ($0 < \alpha_s < 1$)는 시간축 스무딩 변수를 나타낸다. 입력 신호를 주파수-시간축으로 스무딩 해주는 이유는 프레임에서 전력의 최소값 추정시 너무 작은 값으로 추정되는 것을 방지하기 위해서이다. 정의된 윈도우 길이 (D 프레임) 내에서 스무딩된 입력 신호의 최소값이 구해지고, 이를 이용하여 각 주파수 밴드에 대해서 다음과 같이 개략적인 VAD가 수행된다.

$$\gamma_{\min}(k,l) = \frac{|Y(k,l)|^2}{B_{\min} S_{\min}(k,l)}, \zeta(k,l) = \frac{S(k,l)}{B_{\min} S_{\min}(k,l)} \quad (7)$$

$$I(k,l) = \begin{cases} 1 & (\gamma_{\min} < \gamma_0) \text{ and } (\zeta(k,l) < \zeta_0) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

여기서 $S_{\min}(k,l)$ 는 최소값 추정 방식에 의해서 구해진 최소값을 나타내고, B_{\min} 은 구해진 최소 잡음 전력을 보상하기 위한 변수이다. $I(k,l)$ 는 개략적인 VAD 결과로 음성의 부재시 1, 음성의 존재시 0의 값을 갖는다.

식(8)의 결과에 따라서 현재 프레임에서 강한 음성 성분이 다음과 같이 제거 된다.

$$\tilde{S}_f(k,l) = \begin{cases} \frac{\sum_{i=-w}^w b(i) I(k-i,l) |Y(k-1,l)|^2}{\sum_{i=-w}^w b(i) I(k-i,l)}, & \sum_{i=-w}^w I(k-i,l) \neq 0 \\ \tilde{S}(k,l-1), & \text{otherwise} \end{cases} \quad (9)$$

$$\tilde{S}(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s) \tilde{S}_f(k,l) \quad (10)$$

스무딩된 신호 $\tilde{S}(k,l)$ 를 이용하여 정해진 윈도우 내에서 최소값을 추정하고 식(7)과 동일한 방법을 통해서 $\tilde{\gamma}_{\min}(k,l)$, $\tilde{\zeta}(k,l)$ 가 구해지고, 그 결과를 이용하여 사전 음성 부재 확률 $q(k,l)$ 이 구해진다.

$$q(k,l) = \begin{cases} 1 & (\tilde{\gamma}_{\min}(k,l) \leq 1 \\ & \text{and } (\tilde{\zeta}(k,l) < \zeta_0)) \\ (\gamma_1 - \tilde{\gamma}_{\min}(k,l)) / (\gamma_1 - 1) & (1 < \tilde{\gamma}_{\min}(k,l) \\ & \text{and } (\tilde{\zeta}(k,l) < \zeta_0)) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

3. 잡음 환경 분류 기반의 제안된 음성 향상알고리즘

코헨에 의해서 제안된 IMCRA 음성 향상 알고리즘은 강한 음성 성분을 제거하여 보다 정확한 잡음 전력 최소값 추정과 향상된 음성 존재 확률 계산에 의해서 향상된 잡음 전력 추정 성능을 보여주었다 [7]. 하지만 식 (4)에서 볼 수 있는 것처럼 스무딩 파라미터의 값이 $\bar{\alpha}_d = 0.85 \sim 1$ 와 같기 때문에 비정상 잡음처럼 주파수 특성이 빠르게 변화하는 환경에서는 느린 잡음 전력 업데이트에 의해서 음질 저하를 발생할 수 있고, 비교적 느리게 변화하는 정상 잡음의 경우 반대의 이유에 의해서 음질 저하를 발생 시킬 수 있다. 이러한 단점을 보완하기 위해서 본 논문에서는 GMM 기반의 잡음 환경 분류 알고리즘 결과를 기반으로 잡음 전력 추정에 사용되는 스무딩 파라미터가 적응적으로 변화하는 알고리즘을 제안한다 [8]-[10].

GMM은 복 수개의 가우시안 형태의 확률 밀도 함수를 이용하여 각 클래스를 모델링하고, 이를 이용하여 우도비 검정을 통해서 클래스를 분류 하는 패턴 인식기로서 가우시안 확률 밀도함수는 다음과 같이 주어진다.

$$P_n(\vec{x}|\lambda) = \prod_{i=1}^M \alpha_i P_i(\vec{x}), \quad \sum_{i=1}^M \alpha_i = 1. \quad (12)$$

$$P_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T (\Sigma_i)^{-1} (\vec{x} - \vec{\mu}_i)\right\} \quad (13)$$

여기서 $P_n(\vec{x}|\lambda)$ 는 모델 파라미터 $\lambda = \{\alpha_i, \mu_i, \Sigma_i\}$ 조건에서의 조건부확률 값을 나타내고, M 은 가우시안 개수, α 는 각 가우시안 밀도의 가중치, D 는 GMM에 사용되는 특징 벡터의 개수, Σ_i 는 공분산 행렬, μ_i 는 평균 벡터를 나타낸다.

GMM 기반의 패턴 분류기의 성능은 특징벡터의 통계적 편차에 대한 구분이 뚜렷할수록 더욱 우수한 성능을 보여준다. 따라서 잡음 환경 분류 알고리즘을 위한 특징 벡터를 찾기 위해서 스펙트럼 기반의 특징벡터들의 통계적 분류 특성을 조사하였고, 실제로 잡음 환경 (car, white, babble)에 대해서 스펙트럼 센트로이드 (spectrum centroid) 와 스펙트럴 엔트로피 (spectral entropy)의 통계적 분류 특성이 우수하게 나타내었다. 그림 1은 2개의 특징벡터의 통계적 분류 특성을 비교하여 보여주고 있다.

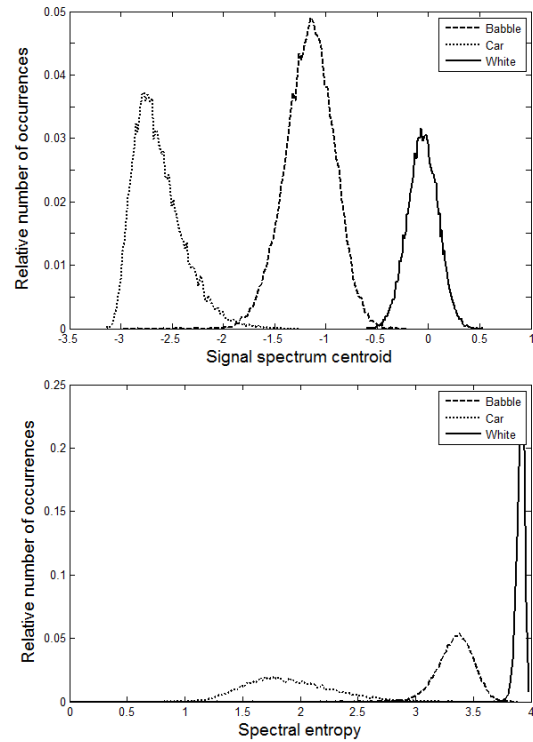


그림 1 GMM 특징벡터 히스토그램
Fig. 1 Histogram of the GMM feature vectors

스펙트럴 센트로이드는 입력된 주파수 스펙트럼의 분포가 1 kHz를 기준으로 해서 얼마나 떨어져서 분포되는지를 나타내는 것으로 다음과 같이 구해진다.

$$SS = \frac{\sum_{k=0}^{(N/2)} \log_2\left(\frac{f(k)}{1000}\right) P(|Y(k,l)|^2)}{\sum_{k=0}^{(N/2)} P(|Y(k,l)|^2)} \quad (14)$$

엔트로피는 확률 변수의 불확실성을 나타내기 위한 최소 정보량 또는 무질서도의 크기를 나타내는 척도로 무질서도가 클수록 높은 엔트로피를 갖는다. 현재 프레임의 주파수 스펙트럼에 대한 엔트로피로 나타내면 다음과 같다.

$$H(|Y(l)|^2) = -\sum_{k=1}^N \{P(|Y(k,l)|^2) \cdot \log_2(P(|Y(k,l)|^2))\} \quad (15)$$

여기서 $P(|Y(k,l)|^2)$ 는 각 주파수 스펙트럼에 대한 확률 값을 나타내고 다음과 같이 계산된다.

$$P(|Y(k,l)|^2) = \frac{|Y(k,l)|^2}{\sum_{k=1}^N |Y(k,l)|^2} \quad (16)$$

일반적으로 잡음 환경은 시간에 따라서 빠르게 변화하지 않기 때문에 매 프레임 간 독립적인 처리 과정에 대한 오류를 보상하고 상관성을 고려하여 스펙트럼 엔트로피 (SE)는 다음과 같이 계산된다.

$$SE(l) = \beta \cdot SE(l-1) + (1-\beta) \cdot H(|Y(l)|^2) \quad (17)$$

여기서 β 는 스무딩 파라미터를 나타낸다. 선택된 특징 벡터를 이용하여 EM (expectation-maximization) 알고리즘 기반의 학습을 통해서 car, white, babble의 모델 $\lambda_c, \lambda_w, \lambda_b$ 를 만들었다. 테스트 과정에서 기화된 잡음 모델에 실제 데이터가 입력되어 구한 우도비 (likelihood ratio)를 계산하여 문턱값과 비교를 통해서 잡음 환경을 분류 한다.

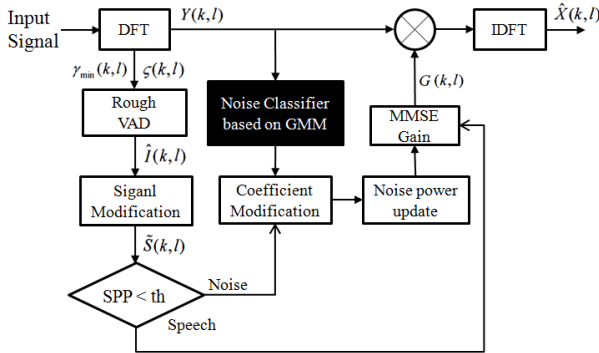


그림 2 제안된 알고리즘의 전체 블록도
 Fig. 2 Block diagram of the proposed method

그림 2는 제안된 알고리즘의 블록다이어그램을 나타낸다. 입력된 신호는 DFT를 통해서 주파수 축 신호로 변환되고, MCRA기반의 VAD를 통해서 강인한 음성 신호가 제거된다. 수정된 입력 신호를 이용하여 정해진 구간 내에서의 최소 전력 값을 추정하고 이를 기반으로 음성 존재 확률이 추출된다. 만약 음성 존재 확률이 기 정의된 문턱 값보다 작을 경우 (즉, 현재 프레임이 음정보다는 잡음에 더 가깝다고 결정될 경우) GMM 기반의 환경 분류 결과에 따라서 잡음 전력 업데이트에 적용되는 스무딩 파라미터가 표 1과 같이 수정되어 잡음 전력이 업데이트 된다.

$$\bar{\lambda}_d(k,l+1) = \bar{\lambda}_d(k,l) \bar{p}(k,l) + [\alpha_d^G \bar{\lambda}_d(k,l) + (1-\alpha_d^G) |Y(k,l)|^2] (1-\bar{p}(k,l)) \quad (18)$$

표 1 잡음 환경 분류 결과에 따른 잡음 전력 추정에 사용되는 스무딩 파라미터 값

Table 1 The smoothing parameter of the noise power

	잡음 환경	스무딩파라미터
α_d^G	car	0.96
	babble	0.55
	white	0.99

추정된 잡음 전력을 이용하여 향상된 음성 신호를 구하기 위한 이득은 Ephraim-Malah에 의해서 제안된 최소 평균 평방 오차 (minimum mean square error, MMSE) 방식을 사용하고, 음성 신호의 추정치는 다음과 같이 나타낸다 [11],[12].

$$\hat{X}(k,l) = G(\gamma(k,l), \xi(k,l)) \cdot Y(k,l) \quad (19)$$

여기서 이득 함수 $G(\gamma(k,l), \xi(k,l))$ 는 다음과 같이 계산된다.

$$G(\gamma(k,l), \xi(k,l)) = \frac{\sqrt{\pi}}{2} \sqrt{\frac{\nu(k,l)}{\gamma^2(k,l)}} \exp\left(-\frac{\nu(k,l)}{2}\right) \times \left[(1+\nu(k,l)) I_0\left(\frac{\nu(k,l)}{2}\right) + \nu(k,l) I_1\left(\frac{\nu(k,l)}{2}\right) \right] \quad (20)$$

$\nu(k,l) = \gamma(k,l)\xi(k,l)/(1+\xi(k,l))$ 를 나타내고, I_0, I_1 는 수정된 베셀 함수의 0차, 1차 버전이다.

4. 실험 및 결과

제안된 음성 향상 알고리즘의 성능을 평가하기 위해서 객관적 음질 평가 방식인 ITU-T P.862 PESQ와 composite measure를 이용하였다. 테스트를 위해서 8 kHz로 샘플링된 8초 길이의 NTT 데이터베이스 (남,녀 각각 100개의 문장)를 사용하였고, 잡음은 NOISEX-92 데이터베이스의 babble, car, white에서 5, 10, 15 dB의 SNR을 갖도록 만들었다. IMCRA에서 사용된 변수는 $\alpha = 0.92, \alpha_d = 0.85, \alpha_s = 0.9, \gamma_0 = 4.6, \zeta_0 = 1.67$ 로 설정하였고, 최소값 추정을 위한 윈도우 길이는 $D=120, U=8, V=15$ 로 설정하였다 [7].

표 2 다양한 잡음 환경에서 기존 IMCRA와 제안된 방법의 PESQ 수치 비교

Table 2 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	2.418	2.750	3.044
Car	IMCRA	3.641	3.888	4.088
	Proposed	3.658	3.903	4.100
White	IMCRA	2.145	2.486	2.839
	Proposed	2.176	2.511	2.860

표 2는 제안된 방법과 기존 IMCRA 방법의 음질 성능 평가를 위해 실시한 PESQ 테스트의 성능을 내고, 테스트 파일에 대한 평균 수치로 나타내었다. 실험 결과, babble, car, white 잡음에 대해 평균적으로 각각 0.033, 0.015, 0.026 정도로 모든 잡음 환경에서 향상된 수치를 보여준다. 특히, 비정상적 잡음 환경인 babble 배경 잡음에서 성능이 크게 향상된 것을 통해서 GMM 기반의 제안된 음성 향상 알고리즘이 비정상 잡음 전력을 효과적으로 추정한다는 것을 알 수 있다.

표 3은 두 시스템의 객관적 성능 평가를 위해 실시한 composite measure 테스트 결과를 나타낸다. composite measure 테스트는 음성 향상된 신호의 음질을 평가하는 PESQ, 정해진 프레임 내에서의 인접한 주파수 밴드 사이의 왜곡 도를 측정하는 weighted-slope spectral distance

표 3 다양한 잡음 환경에서 기존 IMCRA와 제안된 방법의 composite measure 수치 비교

Table 3 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	2.748	3.143	3.473
Car	IMCRA	3.928	4.206	4.436
	Proposed	3.951	4.225	4.451
White	IMCRA	2.336	2.726	3.114
	Proposed	2.379	2.763	3.146

(WSS), 깨끗한 신호와 잡음 처리가 된 신호의 각각에 대해 추출된 LPC를 기반으로 복원된 두 신호의 왜곡 도를 로그 스케일로 측정 한 log-likelihood ratio(LLR)로 구성되어 있고, 식은 다음과 같다 [13].

$$C_{opt} = 1.549 + 0.805PESQ - 0.512LLR - 0.007WSS \quad (21)$$

composite measure에서 사용하는 PESQ는 기존의 PESQ에서 음성의 왜곡과 잡음의 왜곡에 대한 측정치에 가중치를 더 주도록 수정된 측정법을 나타낸다. 실험 결과 babble, car, white 잡음에 대해 평균적으로 각각 0.038, 0.019 0.037로 모든 잡음 환경에서 향상된 수치를 보여준다. 표 3을 통해서 음성의 왜곡도와 인접한 프레임간의 왜곡도 측면에서도 제안된 알고리즘이 모든 잡음 환경에서 더욱 향상된 성능을 보여주는 것을 알 수 있다.

5. 결 론

본 논문에서는 기존 IMCRA 음질 향상 알고리즘의 향상된 잡음 전력 추정을 위해서 잡음 환경에 대한 특성을 고려한 잡음 전력 추정 알고리즘을 제안하였다. 구체적으로 GMM기반의 환경 잡음 분류 알고리즘을 적용하여 정상적인 잡음 환경과 비정상적인 잡음 환경에 대해서 구분하고 그 결과를 기반으로 잡음환경에 적합한 잡음 전력 스무딩 파라미터 값이 적용적으로 변화하도록 하였다. 제안된 음성 향상 기술의 성능을 평가하기 위해서 다양한 잡음 환경에서 PESQ 와 composite measure 테스트를 하였고, 실험 결과 제시된 GMM 기반의 향상된 IMCRA 기법이 기존의 IMCRA 보다 PESQ (babble : 0.033, car : 0.015, white : 0.026), composite measure (babble : 0.038, car : 0.019, white : 0.037)로 향상된 결과를 보여주었다. 특히, 잡음 특성이 시간에 따라서 빠르게 변화하는 비정상 잡음 환경에서 향상된 성능을 나타내었다.

감사의 글

본 연구는 지식경제부 및 정보통신 산업 진흥원의 IT 융합 고급인력과정 지원사업(NIPA-2012-H0401-12-1007)과 서울시 산학연 협력사업 (SS100022) 의 지원에 의하여 이루어진 연구로서, 관계 부처에 감사드립니다.

참 고 문 헌

- [1] R. Martin, "Spectral subtraction based on minimum statistics," Proceeding of 7th EUSIPCO'94, Edinburgh, U.K., pp.1182-1185, Sep. 1994.
- [2] 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.
- [3] Y. H. Son, S. M. Lee, "Improved speech absence probability estimation based on environmental noise classification," Journal of Central South University., Vol. 19, No. 9, pp 2548-2553, September. 2012.
- [4] 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.
- [5] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environment," Signal Processing, pp.2403-2418, Nov. 2001.
- [6] 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.
- [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] P. Renevey and A, Drygajlo, "Entropy based voice activity detection in very noisy conditions." In Eurospeech 2001, pp. 1887-1890, Sep. 2001.
- [10] M. Asgari, A. Sayadian, M. Farhadloo and E. A. Mehrizi, "Voice activity detection using entropy in spectral domain," Telecommunication Networks and Applications Conference 2008, pp. 407-410, Dec. 2008.
- [11] J. Lei, J. Wang and Z, Yang, "A robust voice activity detection algorithm in nonstationary noise," International Conference on Industrial and Information Systems 2009, pp. 195-198, Dec. 2009.
- [12] 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), pp.443-445, Apr. 1985.
- [13] 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), pp.1109-1121, Dec. 1984.

- [13] 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 (宋智鉉)

2007년 2월 인하대학교 전자공학과 학사, 2009년 2월 인하대학교, 전자공학부 석사, 2009년 3월~현재 인하대학교 전자공학부 박사과정

관심분야 : Speech Enhancement, Pattern Classification, Speech Coding.



박 규 석 (朴奎錫)

2010년 2월 인하대학교 전자공학과 학사, 2012년 인하대학교 전자공학과 석사, 2012년~현재 인하대학교 전자공학과 박사 과정

관심분야 : Speech-Signal Processing, Psycho-Acoustic



안 홍 섭 (安洪燮)

2010년 2월 인하대학교 전자공학과 학사, 2012년 인하대학교 전자공학과 석사, 2012년~현재 인하대학교 전자공학과 박사 과정

관심분야 : Bio-Signal Processing, Psycho-Acoustic



이 상 민 (李相旻)

1987년 2월 인하대학교 전자공학과 학사, 1989년 2월 인하대학교 전자공학과 석사, 2000년 인하대학교 전자공학과 박사, 2006년 6월 - 현재 인하대학교 전자공학과 부교수

관심분야 : Brain-Machine interface, Bio-Signal Processing, Psycho-Acoustic