

보청기를 위한 IMCRA 기반 저연산 음성 향상 알고리즘

Low-Complexity Speech Enhancement Algorithm Based on IMCRA Algorithm for Hearing Aids

전유용*, 이상민
Y. Y. Jeon, S. M. Lee

요 약

본 논문에서는 향상된 최소값 제어 재귀 평균 (improved minima controlled recursive averaging, IMCRA) 알고리즘과 로그 최소값 평균 제곱 오차 (log minimum mean square error, logMMSE)를 기반으로 한 저연산 음성 향상 알고리즘을 제안한다. IMCRA 알고리즘은 버퍼를 이용하여 일정 구간에서 입력 신호 전력의 최소값을 추적하고 최소값과 입력 신호의 비율을 통해 음성 존재를 확인한다. 이러한 과정에서 많은 연산이 필요하며 연산량을 줄이기 위해서 음성 존재 확률을 기반으로 한 주파수 밴드별 시변 스무딩으로 최소값을 추적한다. 제안된 알고리즘은 0dB, 5dB, 10dB 그리고 15dB 신호 대 잡음비에서 평균 2.778%, 3.481%, 2.980% 그리고 2.162% 음성 품질이 향상되었으며, 평균 9.570% 연산량이 감소한 것을 확인하였다.

ABSTRACT

In this paper, we proposed a low-complexity speech enhancement algorithm based on a improved minima controlled recursive averaging (IMCRA) and log minimum mean square error (logMMSE). The IMCRA algorithm track the minima value of input power within buffers in local window and identify the speech presence using ratio between input power and its minima value. In this process, many number of operations are required. To reduce the number of operations of IMCRA algorithm, minima value is tracked using time-varying frequency-dependent smoothing based on speech presence probability. The proposed algorithm enhanced speech quality by 2.778%, 3.481%, 2.980% and 2.162% in 0, 5, 10 and 15dB SNR respectively and reduced computational complexity by average 9.570%.

Keyword : Speech Enhancement, Low Complexity Algorithm, Modified IMCRA, Log MMSE, Hearing Aid Algorithm

1. 서론

음성 향상 알고리즘은 입력 채널수, 즉 마이크론의 수에 따라 하나의 마이크론에서 음성을 획

득하여 신호처리를 수행하는 단청 (monaural) 알고리즘과 여러 개의 마이크론에서 음성을 획득하여 처리하는 다중 마이크론 알고리즘으로 나눌 수 있다. 다중 마이크론 기반의 음성 향상 알고리즘이 일반적으로 더 나은 성능을 보이지만 [1], 보청기의 경우 적은 계산량을 요구하기 때문에 주로 단청 음성 향상 알고리즘을 사용한다.

음성 향상 알고리즘은 잡음 환경에서 음성의 품질을 향상 시키는 과정으로 잡음에 의해 오염된 음성에서 잡음을 제거하는 것을 포함한다. 1979년 Boll [2]에 의해서 개발된 스펙트럼 차감법 (spectral subtraction)은 잡음에 의해 오염된 음성에서 잡음 성분을 차감하여 줄이는 방법으로 디지털 신호의 잡음 제거에서 가장 기본적인 알고리즘으로 사용되

접 수 일 : 2017.11.18

심사완료일 : 2017.11.24

게재확정일 : 2017.11.29

* 전유용 : 인하대학교 전자공학과 박사과정

nicejyy@gmail.com (주저자)

이상민 : 인하대학교 전자공학과 교수

sanglee@inha.ac.kr (교신저자)

※ 이 논문은 2017년도 정부(미래창조과학부)의 재원으로 산업기술연구회의 지원을 받아 수행된 한국전기연구원(KERI) 주요사업입 (No. 17-12-N0101 -46)

※ 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업입(No. NRF-2016R1A2B4015370)

고 있다. 스펙트럼 차감법 등의 알고리즘을 이용하여 잡음을 제거하기 위해 가장 중요하게 선행되어야 하는 과정이 잡음의 전력을 추정하는 것이다. 잡음 전력을 추정하기 위한 전통적인 방법 중 하나는 음성 검출기 (voice activity detection, VAD)를 이용하여 음성 부재 구간에서 잡음의 평균을 구하는 것이다[2-4]. 하지만 이는 음성 검출기의 성능에 의존하며 신호 대 잡음비 (signal to noise ratio, SNR)가 낮은 구간에서는 음성 검출 성능이 떨어지기 때문에 왜곡된 음성을 출력하는 단점을 갖고 있다. 이를 개선하기 위해서 음성 검출기를 사용하지 않고 스무딩 (smoothing)을 이용하여 탐색 윈도우 안에서 최소 전력을 계산하고 스무딩 된 스펙트럼 전력과 비를 이용하여 음성 존재를 파악하고 사후 신호 대 잡음비 (a posteriori SNR)를 이용하여 잡음을 제거하는 방법인 최소값 제어 재귀 평균 (minima controlled recursive averaging, MCRA) 알고리즘이 제안되었다 [5]. 이 알고리즘은 음성 신호가 잡음 신호에 비해 변화가 많다는 것과 각 주파수 밴드 사이의 전력은 독립적이라는 것을 가정하고 있다. 또한 로컬 윈도우 내의 최소값과 현재 전력의 비를 문턱값과 비교하여 0과 1로 음성 존재를 결정한 후에, 1차 재귀 방법을 통해 음성 존재 확률을 결정한다. 이러한 특성 때문에 음성과 같이 변화가 많은 다중 누화 잡음 (babble noise)에서는 효과가 떨어지며, 각각의 주파수 밴드를 독립적으로 보기 때문에 잡음 제거 후에 주파수 밴드별 전력 차이가 크게 발생하는 문제점이 있었다. 2003년 Cohen [6]은 최소값 제어 재귀 평균 알고리즘의 문제점을 개선하기 위해서 향상된 최소값 제어 재귀 평균 (improved minima controlled recursive averaging, IMCRA) 알고리즘을 제안하였다. 개선된 최소값 제어 재귀 평균 알고리즘에서는 이전 알고리즘과 마찬가지로 입력 신호의 전력과 로컬 윈도우에서의 그 최소값을 비교하여 음성 존재 확률을 결정하지만 0과 1만으로 경판정 (hard decision)과 음성 존재 확률을 계산하는 연판정 (soft decision)을 이용한다는 것이 다른 점이라고 할 수 있다. 또한 각 주파수 밴드 사이를 완전히 독립적인 것이 아닌 어느 정도의 연관성이 있는 것으로 판단하여 잡음 전력을 추정한다.

보청기와 같이 연산이 제한된 경우에는 알고리즘들의 연산량을 줄여 효율을 높이는 것과 잡음 성분을 효율적으로 추적하는 것이 중요하다. 그러한 관점에서 볼 때, IMCRA 알고리즘은 몇 가지 단점을 갖는다. 첫 번째 단점은 음성 존재 확률을 추정하기 위한 최소값을 탐색하기 위해서 버퍼 (buffer)

가 필요하며, 일정 프레임 (frame)마다 버퍼 내에서 최소값을 찾기 위해서는 버퍼 크기에 비례하는 최소값 연산을 필요로 한다는 것이다. 두 번째로 버퍼 내에서 최소값을 추출하기 때문에 변화가 큰 잡음에서 잡음을 정확하게 추정하기 어렵다는 단점이 있다.

본 논문에서는 IMCRA 알고리즘의 최소값 추적 방법의 단점을 개선하여 효율적인 잡음 추적이 가능한 저연산 IMCRA (low-complexity IMCRA, LIMCRA)를 제안하며, 추정된 잡음을 제거하기 위해서 로그 최소값 평균 제곱 오차 (logMMSE) 알고리즘 [7]을 사용하였다. 알고리즘의 성능을 평가하기 위해서 기존 잡음 데이터베이스 뿐만 아니라 시간에 따른 변화가 큰 실제 잡음 상황을 녹음하여 사용하였다. 제안된 알고리즘의 성능을 평가하기 위해서 기존 잡음 데이터베이스에서 5가지 잡음을 선택하고 주변에서 5가지 잡음을 녹음하였으며, 문장인지 검사를 위해 개발된 한국어 문장을 녹음하여 사용하였다.

2장에서 IMCRA 잡음 추정 알고리즘을 리뷰하고 단점을 분석하며, 3장에서는 단점을 극복하기 위한 잡음 추정 알고리즘을 제안한다. 4장에서 제안된 알고리즘의 평가를 위한 데이터에 대한 설명과 평가 방법을 설명하며, 5장에서 결론을 제시한다.

2. IMCRA 알고리즘

2.1 IMCRA 알고리즘 리뷰

시간 영역에서의 잡음에 의해 오염된 음성 신호 (noisy signal, 노이지 신호)를 y 라고 표현할 때, y 는 깨끗한 음성 신호 x 와 음성 신호와 상관관계가 없는 잡음 d 의 합이라고 할 수 있다. 노이지 신호에서 잡음 신호를 추정하기 위해서 일반적으로 잡음의 유무에 따라서 두 가지 가설로 나눌 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 $Y(k,l)$, $X(k,l)$, 그리고 $D(k,l)$ 는 l 번째 프레임에서 k 번째 주파수 밴드에서의 입력된 노이지 신호, 깨끗한 음성 신호, 그리고 잡음 신호의 스펙트럼 (spectrum)이다.

잡음 추정 알고리즘을 비롯해 시간에 따라 상관관계를 갖는 신호에 대한 신호처리에서 가장 기본적인 것 중 하나는 입력 신호에 대한 스무딩을 통해 시간 변화에 따른 상관관계를 이용하여 출력

추정하는 것이다. 잡음 추정 알고리즘에서 또한 식 (2)과 같이 과거에 추정한 잡음 전력과 입력된 신호의 전력 사이에서 스무딩을 통해 현재의 잡음 전력을 추정한다.

$$\bar{\lambda}_d(k,l) = \tilde{\alpha}_d(k,l)\bar{\lambda}_d(k,l-1) + (1-\tilde{\alpha}_d(k,l))|Y(k,l)|^2 \quad (2)$$

여기서 $\bar{\lambda}_d(k,l)$ 는 추정된 잡음 전력을 의미하며, $\tilde{\alpha}_d$ 는 주파수 밴드별 시변 스무딩 파라미터로 다음과 같이 고정된 스무딩 파라미터인 α_d 와 조건부 음성 존재 확률 $p(k,l)$ 의해 조절된다.

$$\tilde{\alpha}_d = \alpha_d + (1-\alpha_d)p(k,l) \quad (3)$$

조건부 음성 존재 확률은 입력 신호의 전력 스펙트럼이 주어졌을 때, 음성이 존재할 확률을 의미하여 식 (4)와 같이 계산할 수 있다.

$$p(k,l) = P(H_1(k,l) | |Y(k,l)|^2) = \left\{ 1 + \frac{q(k,l)}{1-q(k,l)} (1+\xi(k,l)) \exp(-\nu(k,l)) \right\}^{-1} \quad (4)$$

여기서 $q(k,l)$ 는 사전 음성 부재 확률 (a priori probability for speech absence), $P(H_0(k,l))$ 이고 $\nu(k,l) = \gamma(k,l)\xi(k,l)/(1+\xi(k,l))$, 여기서 사후 신호 대 잡음비 (a posteriori SNR)와 사전 신호 대 잡음비 (a priori SNR)는 $\gamma(k,l) = |Y(k,l)|^2/\lambda_d(k,l)$, $\xi(k,l) = \lambda_x(k,l)/\lambda_d(k,l)$ 와 같이 정의되며, λ_d 와 λ_x 는 각각 잡음과 음성의 전력을 의미한다.

사후 신호 대 잡음비는 음성 부재 구간에서 획득한 잡음 전력을 이용하여 직접 추정할 수 있지만 사전 신호 대 잡음비는 깨끗한 음성 신호의 전력을 필요로 하지만 직접적으로 음성만 있는 경우는 가정하지 않으며, 음성만 있는 경우라고 하더라도 음성은 잡음에 비해 변화가 많기 때문에 직접적으로 획득하는데 한계가 있다. 그러한 이유로 사전 신호 대 잡음비는 decision directed (DD) 기법을 이용해 다음 같이 추정한다 [8].

$$\hat{\xi}(k,l) = \alpha \frac{|\hat{X}(k,l-1)|^2}{\lambda_d(k,l-1)} + (1-\alpha)\max(\gamma(k,l)-1,0) = \alpha \frac{|Y(k,l-1)*G(k,l-1)|^2}{\lambda_d(k,l-1)} + (1-\alpha)\max(\gamma(k,l)-1,0)$$

$$= \alpha G^2(k,l-1)*\gamma(k,l-1) + (1-\alpha)\max(\gamma(k,l)-1,0) \quad (5)$$

여기서 G 는 입력신호에서 음성 신호를 추출하기 위한 주파수 밴드별 이득을 의미하며, α 는 스무딩 파라미터를 의미한다.

식 (4)의 조건부 음성 존재 확률을 계산하기 위해서 사전 음성 부재 확률, $q(k,l)$ 를 계산해야 한다. IMCRA 알고리즘에서 사전 음성 부재 확률을 계산하기 위해 다음과 같은 과정을 거친다. 연속된 프레임에서 음성의 강한 상관관계를 고려하고, 주파수 밴드 사이에서의 상관관계를 고려하기 위해 다음과 같이 입력된 신호를 주파수 축과 시간 축에서 스무딩을 수행한다.

$$S_f(k,l) = \sum_{i=-w}^w b(i) |Y(k-i,l)|^2 \quad (6)$$

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)S_f(k,l) \quad (7)$$

여기서 $b(i)$ 는 길이가 $2w+1$ 인 정규화 된 해닝창 (Hanning window)을 나타내며, $\alpha_s (0 < \alpha_s < 1)$ 은 스무딩 파라미터를 의미한다.

유한한 지역 윈도우 (local window)에서 스무딩된 스펙트럼의 최저값, 을 이용해서 다음과 같이 음성 유무를 대략적으로 추출한다.

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

여기서 γ_0 와 ζ_0 는 음성 유무를 판단하기 위한 문턱값 (threshold)이며, $\gamma_{\min}(k,l)$ 와 $\zeta(k,l)$ 는 다음과 같이 계산한다.

$$\gamma_{\min}(k,l) = \frac{|Y(k,l)|^2}{B_{\min} S_{\min}(k,l)}, \quad \zeta(k,l) = \frac{S(k,l)}{B_{\min} S_{\min}(k,l)} \quad (9)$$

여기서 $S_{\min}(k,l)$ 은 일정한 범위의 프레임에서 $S(k,l)$ 의 최소값을 의미하며, B_{\min} 은 최소값과 실제 잡음 전력의 차이를 보상하기 위한 바이어스 (bias)로서, 입력 신호의 종류, 데이터의 길이, 오버랩의 길이, 분석창의 길이 등에 따라서 달라질 수 있지만 보통의 경우는 고정해서 사용한다.

식 (8)에 의해 잡음으로 판별된 요소들의 스펙트럼만을 이용하여 다음과 같이 스무딩을 진행한다.

$$\tilde{S}_f(k,l) = \begin{cases} \frac{\sum_{i=-w}^w b(i)I(k-i,l) |Y(k-i,l)|^2}{\sum_{i=-w}^w b(i)I(k-i,l)} & \text{if } \sum_{i=-w}^w I(k-i,l) \neq 0 \\ \tilde{S}(k,l-1) & \text{otherwise} \end{cases} \quad (10)$$

$$\tilde{S}(k,l) = \alpha_s \tilde{S}(k,l-1) + (1-\alpha_s) \tilde{S}_f(k,l) \quad (11)$$

첫 번째 스무딩 후, 식 (8)에서 음성 유무를 판별한 것과 같이 두 번째 스무딩 후에도 음성 유무, 즉 사전 음성 부재 확률을 다음과 같이 계산한다.

$$q(k,l) = \begin{cases} 1 & \text{if } (\tilde{\gamma}_{\min}(k,l) \leq 1) \\ & \text{and } (\tilde{\zeta}(k,l) < \zeta_0) \\ \frac{\gamma_1 - \tilde{\gamma}_{\min}(k,l)}{\gamma_1 - 1} & \text{if } (1 < \tilde{\gamma}_{\min}(k,l) < \gamma_1) \\ & \text{and } (\tilde{\zeta}(k,l) < \zeta_0) \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

여기서 γ_1 와 ζ_0 는 문턱값이며, $\tilde{\gamma}_{\min}(k,l)$ 과 $\tilde{\zeta}(k,l)$ 는 식 (9)와 같이 계산되며 S_{\min} 대신 \tilde{S} 의 구간 최소값인 \tilde{S}_{\min} 을 사용한다.

이러한 과정을 통해 식 (4)의 조건부 음성 존재 확률을 계산할 수 있으며, 식 (1)과 식 (2)를 이용해서 잡음을 추정할 경우 추정된 잡음의 크기가 실제 잡음의 크기보다 작게 추정되기 때문에 다음과 같이 보정한다.

$$\bar{\lambda}_d(k,l) = \beta \times \lambda_d(k,l) \quad (13)$$

2.2 IMCRA 알고리즘 단점

IMCRA 알고리즘을 이용한 잡음 추정에는 몇 가지 단점이 있다. 첫 번째로, 2.1장에서 언급한 것과 같이 IMCRA 알고리즘은 두 가지의 최소값을 추적한다.

$$\begin{aligned} S_{\min}(k,l) &= \min \{S(k,l') | l-D+1 \leq l' \leq l\} \\ \tilde{S}_{\min}(k,l) &= \min \{\tilde{S}(k,l') | l-D+1 \leq l' \leq l\} \end{aligned} \quad (14)$$

여기서 D 는 최소값을 탐색하는 윈도우의 크기를 그리고 l 은 프레임 색인을 의미한다.

과거 D 프레임에서의 최소값을 구하기 위해 $V(=15)$ 프레임 동안 2개의 최소값 S_{\min} , \tilde{S}_{\min} 과, 2개의 임시 최소값 S_{\min_sw} , \tilde{S}_{\min_sw} 을 구한다.

$$\begin{aligned} S_{\min}(k,l) &= \min \{S_{\min}(k,l-1), S(k,l)\} \\ \tilde{S}_{\min}(k,l) &= \min \{\tilde{S}_{\min}(k,l-1), \tilde{S}(k,l)\} \\ S_{\min_sw}(k,l) &= \min \{S_{\min_sw}(k,l-1), S(k,l)\} \\ \tilde{S}_{\min_sw}(k,l) &= \min \{\tilde{S}_{\min_sw}(k,l-1), \tilde{S}(k,l)\} \end{aligned} \quad (15)$$

또한 V 프레임마다 임시 최소값 S_{\min_sw} 와 \tilde{S}_{\min_sw} 을 크기가 $N(=8)$ 인 버퍼에 저장하며, V 프레임마다 S_{\min} 과 \tilde{S}_{\min} 을 버퍼 내에서 각 주파수 밴드의 최소값으로 업데이트하고, 임시 최소값을 각각 $S(k,l)$ 과 $\tilde{S}(k,l)$ 으로 업데이트한다. 이러한 과정은 쉽게 수행이 가능하지만 데이터 복사와 데이터 값 비교와 같은 간단한 연산을 많이 요구한다. 만약 버퍼의 사이즈가 N , 주파수 분석 길이를 M 이라고 가정할 때, 매 프레임마다 2가지 최소값과 2가지 임시 최소값을 계산하는데 $(M/2+1) \times 4$ 번의 비교 연산이 필요하고, V 프레임마다 최소값과 업데이트하기 위해서 $(M/2+1) \times (N-1) \times 2$ 비교가 필요하며, 임시 최소값의 저장과 업데이트를 위해 $(M/2+1) \times 4$ 번의 데이터 복사가 필요하다.

또한 최소값을 추정하여 잡음 전력의 개념으로 사용하고 있기 때문에 잡음의 전력이 증가하는 경우, 최소 에너지는 잡음 전력의 증가에 따라서 증가하는 모습을 보여야 한다. 하지만 V 프레임마다 버퍼에 최소값을 저장하는 방법으로 장시간 동안의 최소값을 최종 최소값으로 결정하는 과정을 거치기 때문에 잡음의 변화가 큰 경우에는 잡음을 빠르게 추적할 수 없다는 것이 두 번째 단점이라고 할 수 있다. 만약 버퍼의 길이를 줄여서 빠르게 추적할 수는 있지만, 음성에 있는 경우에도 빠른 추적을 통해서 음성이 잡음으로 추정되는 경우가 발생할 수 있다.

3. 제안된 지연산 음성 향상 알고리즘

본 장에서는 IMCRA 알고리즘의 단점을 극복하기 위한 잡음 추정 알고리즘을 제안하며, 로그 최소값 평균 제곱 오차 알고리즘을 이용한 잡음 제거에 대하여 설명한다. 그림 1은 본 논문에서 제안한 지연산 음성 향상 알고리즘의 블록도이다. 제안된 알고리즘은 크게 잡음 추정 알고리즘과 잡음 제거 알고리즘으로 구성된다. 기존 잡음 추정 알고리즘은 최소값을 찾기 위해서 버퍼에 데이터를 복사하고 버퍼 내에서 최소값을 검색하기 위해서 많은 연산을 필요로 하며, 변화가 많은 잡음에서는 잡음 전력을 추적하는데 한계가 있다. 이를 개선하기 위해서 버퍼 내에서 최소값을 추적하는 대신 음성 존재 확률을 이용한

간단한 스무딩 방식을 이용하여 최소값을 추적하도록 하였으며, 추정된 잡음을 제거하기 위해서 로그 최소값 평균 제곱 오차 알고리즘을 사용하였다.

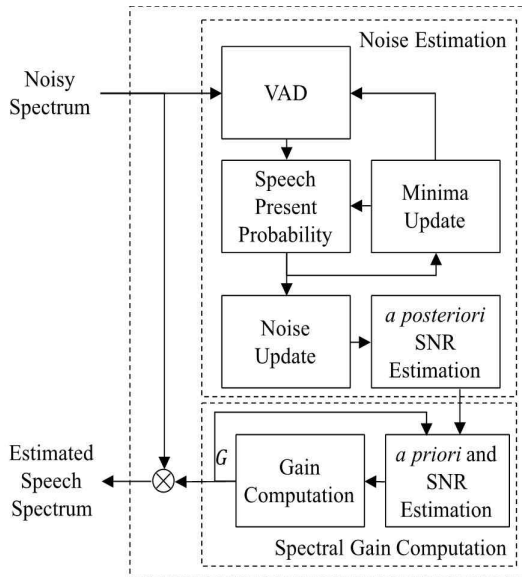


그림 1. 제안된 저연산 음성 향상 알고리즘의 블록도
Fig. 1. Block diagram of proposed low-complexity speech enhancement algorithm

대부분의 잡음 추정 알고리즘에서는 잡음의 스펙트럼이 음성에 비해서 덜 변한다고 가정한다. 이 가정에 따라서, 1에 가까운 스무딩 파라미터를 이용하여 스무딩 된 스펙트럼의 최소 전력은 잡음의 스펙트럼으로 간주된다. 또한 식(9)에서 보면 최소값은 사후 신호 대 잡음비를 도출하는데 사용한다. 그렇기 때문에 최소값은 잡음의 스펙트럼과 매우 관련이 있으며, 잡음 스펙트럼을 고려하여 결정되어야 한다. 이를 위해 잡음 존재 확률을 확인하는 것이 필요하며, 식 (4)의 조건부 음성 존재 확률 (conditional speech presence probability)을 이용한다. IMCRA 알고리즘은 식 (1)과 식 (2)에 따라서 음성 존재 확률을 기반으로 입력 신호의 전력 중 잡음 전력이 어느 정도 인지를 결정한다. 이러한 잡음 전력을 추정 방식을 모방하여 최소값을 추적하는 방법을 제안한다.

그림 2는 제안된 저연산 잡음 추정 알고리즘의 블록도이다. 제안된 잡음 추정 알고리즘에서는 기존 IMCRA 알고리즘과 마찬가지로, 매 프레임마다 최소값을 계산하되, $U(=30)$ 프레임 마다 음성 존재 확률을 적용하여 다음과 같이 최소값을 업데이트 한다.

$$\begin{aligned} S_{min}(k,l) &= \alpha_m S_{min}(k,l) + (1-\alpha_m)S(k,l) \\ \tilde{S}_{min}(k,l) &= \tilde{\alpha}_n(k,l)\tilde{S}_{min}(k,l) + (1-\tilde{\alpha}_n(k,l))\tilde{S}(k,l) \end{aligned} \quad (16)$$

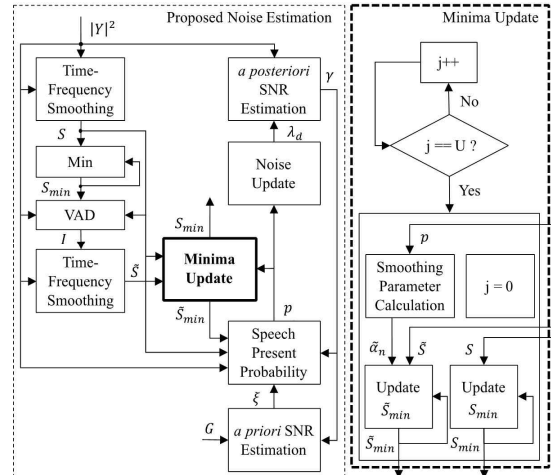


그림 2. 제안된 저연산 잡음 추정 알고리즘의 블록도
Fig. 2. Block diagram of proposed low-complexity noise estimation algorithm

여기서 α_m 은 고정된 스무딩 파라미터이며, 입력 에너지와 최소값의 차이를 고려하여 0.99를 사용하며, $\tilde{\alpha}_n$ 은 주파수 밴드별 시변 스무딩 파라미터이며 다음과 같이 음성 존재 확률에 의해서 조절된다.

$$\tilde{\alpha}_n(k,l) = \alpha_n + (1-\alpha_n)p(k,l) \quad (17)$$

여기서 α_n 은 최소값 업데이트 정도를 결정하는 스무딩 파라미터로 음성이 없는 경우 빠른 에너지 업데이트를 위해 0.6을 사용하였다.

최소값을 업데이트 하는 과정을 보면 $S_{min}(k,l)$ 은 고정된 스무딩 파라미터를 사용하고 $\tilde{S}_{min}(k,l)$ 은 음성 존재 확률에 따라 변화하는 주파수 밴드별 파라미터를 사용한다. $S_{min}(k,l)$ 은 입력 신호에서 1차 음성 검출기에 사용되는데 이후에 시간-주파수 스무딩을 수행하기 때문에 성능에 결정적인 역할을 하지 않는다고 판단되므로 고정된 값을 사용하였고, $\tilde{S}_{min}(k,l)$ 의 경우는 음성 부재 확률을 계산하는 기준에 영향을 줄 수 있고, 이는 조건부 음성 존재 확률에 영향을 미쳐 전체 음질에 영향을 줄 수 있기 때문에 음성 존재 확률을 사용한 주파수 밴드별 시변 스무딩 파라미터 업데이트를 사용하였다.

제안된 알고리즘에 의해 추정된 잡음 전력과 입력 신호의 전력을 이용하여 사후 신호 대 잡음비를 업데이트하며, 식 (5)를 이용하여 사전 신호 대 잡음비를 업데이트 하고, 다음을 이용하여 각 주파수 밴드의 이득을 계산한다.

$$G(k,l) = \frac{\xi(k,l)}{1+\xi(k,l)} \exp\left(\frac{1}{2} \int_{\nu(k,l)}^{\infty} \frac{e^{-t}}{t} dt\right) \quad (18)$$

여기서 $\nu(k,l)$ 은 식 (4)에서 사용하는 것과 같이 계산할 수 있다. 이러한 방법은 이득 계산 방법은 좋은 성능을 보이지만 [9] 무한대의 범위에서의 지수 적분 (exponential integral) 필요하기 때문에 저연산에는 적합하지 않을 수 있다. 이를 해결하기 위해서 일정 범위 내에서는 같은 이득을 내도록 테이블을 만들어 사용하였다.

4. 성능 평가 및 고찰

알고리즘의 성능을 평가하기 위해서 NoiseX-92 데이터베이스에서 5가지 잡음 (babble, factory1, tank, volvo, white)을 선택하였고, 데이터베이스 보다 더 변화가 큰 5가지 실제 잡음 환경 (cafe, laundry, restaurant, subway, traffic)에서 잡음을 녹음하고 10가지 잡음을 16kHz로 샘플링하였다. 또한 한국어 음성에 대한 테스트를 위해 문장 인지 검사를 위해 개발된 문장[10] 중 10가지 문장을 남성이 발성하도록 하고 녹음하여 사용하였다. 이렇게 녹음된 잡음과 문장은 0dB, 5dB, 10dB, 그리고 15dB 신호 대 잡음비로 섞어 사용하였다.

먼저 기존 알고리즘과 제안된 알고리즘의 연산량을 측정하였다. 알고리즘의 전체 연산량은 주파수 밴드와 알고리즘의 복잡도에 따라 달라질 수 있다. 그렇기 때문에 주파수 밴드(=129)를 동일하게 설정하였으며, 한 프레임에서 한 주파수 성분에 대한 연산량을 계산하였다. 연산량은 덧셈, 뺄셈, 곱셈, 나눗셈, 지수승, 로그 등의 연산을 각각 1회만에 수행할 수 있다고 가정하였다. 기존 잡음 추정 알고리즘에서는 평균 62.401 번의 연산이 필요하였으며, 제안된 알고리즘에서는 평균 56.429 번의 연산이 필요하여, 기존 알고리즘에 비해 9.570% 줄어든 것을 확인하였다. 이 결과는 한 프레임의 전체 주파수 밴드에서 잡음을 추정하는데 770.388 번의 연산을 줄일 수 있는 것을 의미한다.

또한 각 신호 대 잡음비로 섞인 신호를 각 알고리즘에 통과한 후 다음과 같이 출력된 음성의 품질을 측정하였다 [11].

$$C_{ovl} = \beta_0 + \beta_1 IS + \beta_2 PESQ \tag{19}$$

여기서 IS 는 Itakura-Saito 거리를 의미하며, $PESQ$ 는 인지적 음성 품질 평가 (perceptual evaluation of the speech quality)를 의미하며, β_0 , β_1 , 그리고 β_2 는 각 평가에 대한 가중치로 각각 -0.736, -0.012 그리고 1.5를 사용한다.

표 1은 각 잡음 환경에서 입력 신호 (noisy), 기존 알고리즘을 사용한 경우 (IMCRA)와 제안된 잡음 추정 알고리즘을 사용한 경우 (proposed)의 음성 품질을 나타낸 것이다. 표에서 볼 수 있듯이, 대부분의 조건에서 제안된 알고리즘에 의한 음성 품질이 기존 알고리즘에 의한 음성 품질보다 향상된 것을 볼 수 있다. 0dB laundry 잡음 환경에서 기존 알고리즘에 비해서 음성 품질이 저하된 것을 볼 수 있지만, 그 차이가 1% 미만이며, 각 신호 대 잡음비에서 평균 2.778%, 3.481%, 2.980% 그리고 2.162% 향상된 것을 볼 수 있다. 또한 제안된 알고리즘이 babble, tank, cafe, restaurant, 그리고 subway와 같이 변화가 큰 잡음에서 더 좋은 성능을 보이는 것으로 나타났다.

표 1. 노이지 신호, IMCRA와 제안된 알고리즘에 의해 향상된 음성 신호의 전체 음성 품질
Table 1. Overall speech quality of noisy signal, enhanced speech by IMCRA and proposed algorithm.

SNR	noise	proposed	IMCRA	noisy
0dB	babble	1.095	1.030	0.861
	factory1	0.917	0.901	0.755
	tank	1.302	1.236	0.828
	volvo	2.103	1.948	1.465
	white	0.928	0.924	0.709
	cafe	0.745	0.731	0.704
	laundry	0.770	0.773	0.705
	restaurant	0.702	0.694	0.654
	subway	0.957	0.933	0.821
	traffic	0.722	0.717	0.626
5dB	babble	1.463	1.394	1.090
	factory1	1.201	1.158	0.896
	tank	1.698	1.643	1.093
	volvo	2.631	2.348	2.065
	white	1.247	1.241	0.786
	cafe	0.885	0.877	0.807
	laundry	0.952	0.942	0.794
	restaurant	0.784	0.769	0.700
	subway	1.382	1.329	0.990
	traffic	0.971	0.950	0.671
10dB	babble	1.902	1.850	1.513
	factory1	1.665	1.621	1.217
	tank	2.168	2.051	1.574
	volvo	3.228	3.047	2.747
	white	1.720	1.716	0.984
	cafe	1.175	1.163	1.017
	laundry	1.292	1.268	1.001
	restaurant	0.998	0.957	0.839
	subway	2.047	1.990	1.298
	traffic	1.391	1.359	0.766
15dB	babble	2.381	2.342	2.153
	factory1	2.202	2.178	1.840
	tank	2.797	2.701	2.242
	volvo	3.597	3.491	3.585
	white	2.269	2.268	1.397
	cafe	1.648	1.633	1.408
	laundry	1.832	1.823	1.441
	restaurant	1.360	1.301	1.108
	subway	2.827	2.722	1.852
	traffic	1.985	1.937	0.991

그림 3은 0dB SNR babble 환경에서의 시간에 따른 실제 잡음 전력 (회색 실선), IMCRA 알고리즘에 의해 추정된 잡음 전력 (검은색 점선), 그리고 제안된 알고리즘에 의해 추정된 잡음 전력 (검은색 실선)을 나타낸다. 제안된 알고리즘에 의해 추정된 잡음의 전력이 기존 잡음 추정 알고리즘에 의해 추정된 잡음 전력 보다 실제 잡음의 전력을 더 잘 추적하는 것을 볼 수 있다.

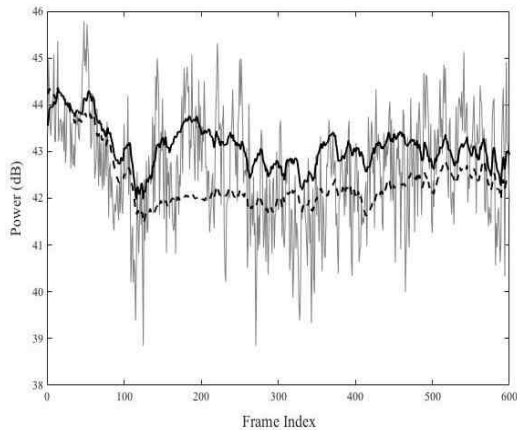


그림 3. 0dB babble 환경에서 평균 잡음 전력 (회색 실선), IMCRA 알고리즘에 의해 추정된 잡음 전력 (검은색 점선), 제안된 알고리즘에 의해 추정된 잡음 전력 (검은색 실선)

Fig. 3. Noise power averaged in all frequency at 0dB babble noise. Noise power (gray solid line), estimated noise power by IMCRA (black dotted line), and proposed algorithm (black solid line)

5. 결론

여러 가지 이유로 단청 알고리즘이 필요한 여러 가지 경우가 있다. 보청기에서 안전상의 이유로 하드웨어의 성능을 제한하고 있어서 저연산 알고리즘이 가장 많이 필요한 분야 중 하나이다. 본 논문에서는 보청기를 위한 저연산 잡음 추정 알고리즘을 제안하였고, 기존 알고리즘에 비해서 9.570% 연산량을 줄일 수 있었고, 0dB, 5dB, 10dB, 그리고 15dB 신호 대 잡음비에서 2.778%, 3.481%, 2.980% 그리고 2.162% 음성 품질이 향상되는 것을 확인하였다. 0dB laundry 잡음 환경에서 음성 품질이 저하되었지만 0.374%로 매우 작은 수준이며 이외의 모든 환경에서는 향상된 것을 볼 수 있었고, 특히 babble, tank, cafe, restaurant, 그리고 subway같이 변화가 큰 잡음에서 기존 잡음에 비해서 더 좋은 성능을 보이는 것을 볼 수 있었다.

REFERENCES

- [1] Kochkin, Sergei. "Customer satisfaction with single and multiple microphone digital hearing aids." *Hearing Review* 11 pp. 24-34, 2000.
- [2] Boll, Steven. "Suppression of acoustic noise in speech using spectral subtraction." *IEEE Transactions on acoustics, speech, and signal processing* 27.2 pp. 113-120, 1979.
- [3] Ephraim, Yariv, and David Malah. "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 32.6 pp. 1109-1121, 1984.
- [4] Ephraim, Yariv, and David Malah. "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33.2 pp. 443-445, 1985.
- [5] Cohen, Israel, and Baruch Berdugo. "Noise estimation by minima controlled recursive averaging for robust speech enhancement." *IEEE signal processing letters* 9.1 pp. 12-15, 2002.
- [6] Cohen, Israel. "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging." *IEEE Transactions on speech and audio processing*, Vol.11, Issue 5 pp. 466-475, 2003. DOI:10.1109/TSA.2003.811544.
- [7] Ephraim, Yariv, and David Malah. "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33.2 pp. 443-445, 1985.
- [8] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech, Sig. Process.*, vol. ASSP-32, no. 6, pp. 1190-1121, Dec. 1984.
- [9] Hu, Yi, and Philipos C. Loizou. "Subjective comparison of speech enhancement algorithms." *Acoustics, Speech and Signal Processing*, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. vol. 1, IEEE, 2006.
- [10] Jang, H. S., et al. "Development of Korean standard sentence lists for sentence recognition tests." *Audiology* 4.2 pp. 161-177, 2008.

- [11] Hu, Yi, and Philipos C. Loizou. "Evaluation of objective quality measures for speech enhancement." IEEE Transactions on audio, speech, and language processing 16.1 pp. 229-238, 2008.



전 유 용 (Yuyong Jeon)

2009년 02월 인하대학교 전자 공학과 석사 졸업
2009년 03월 - 현재 인하대학교 전자 공학과 박사과정

Interest: Speech Enhancement, Noise Reduction, Signal Processing



이 상 민 (Sangmin Lee)

1987년 2월 인하대학교 전자 공학과 학사 졸업
1989년 2월 인하대학교 전자 공학과 석사 졸업.
2000년 2월 인하대학교 전자 공학과 박사 졸업
2006년 6월 - 현재 인하대학교 전자공학과 교수

Interest: Bio-signal processing, PsychoAcoustic, Brain-Machine Interface