

음성활동영역검색을 사용하는 유색잡음에 오염된 음성의 향상을 위한 일반화 부공간 접근

손경식¹ · 김현태^{2*}

A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise Using Voice Activity Detector(VAD)

Kyung-sik Son¹ · Hyun-tae Kim^{2*}

¹ Department of Electronics Engineering, Pusan National University, Busan 609-735, Korea

² Department of Multimedia Engineering, Dongeui University, Busan 614-714, Korea

요 약

본 논문에서는 유색잡음에 의해 오염된 음성신호의 음성향상 알고리즘인 YL 접근법에 VAD(voice activity detector)를 구현하는 수정된 알고리즘을 제안한다. 제안한 알고리즘을 YL 접근법 및 LS 접근법과 컴퓨터 시뮬레이션으로 성능을 비교하였다. 사용한 유색잡음은 자동차 잡음과 다중화자 배블 잡음으로 AURORA 데이터베이스로부터 각각 발췌하였고, 음성신호는 TIMIT 데이터 베이스로부터 발췌하였다. 제안한 알고리즘을 실험했을 때 제안하는 방법이 신호대잡음비 및 스펙트럼 왜곡 측면에서 기존의 두 알고리즘 보다 개선됨을 확인하였다.

ABSTRACT

In this paper, we proposed the modified YL(Yi and Loizou) algorithm, using a VAD(voice activity detector) for enhancing speech corrupted by colored noise. The performance of the proposed algorithm has been compared to the YL algorithm and LS(Lee and Son, etc.) algorithm by computer simulation. The colored noises used in the experiment were a car noise and multi-talker babble from the AURORA data base and the used voices from the TIMIT data base. It is confirmed that the proposed algorithm shows better performance from SNR(signal to noise ratio) and SSD(speech spectral distortion) viewpoint over the previous two approach.

키워드 : 음성 향상, 일반화 부공간 접근, 음성활동영역 검색

Key word : Speech Enhancement, Generalized Subspace Approach, Voice Activity Detector

접수일자 : 2013. 04. 09 심사완료일자 : 2013. 05. 03 게재확정일자 : 2013. 05. 22

* **Corresponding Author** Hyun-Tae Kim(E-mail:htaekim@deu.ac.kr, Tel:+82-51-890-1992)

Department of Multimedia Engineering, Dongeui University, Busan 614-714, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2013.17.8.1769>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

음성 향상을 위한 부공간에 근거하는 새로운 접근법이 Epharaim 및 Van trees(EV)[1]에 의해 제안되었다. EV는 미리 정해진 임계치 이하로 잔여잡음을 유지하면서 음성왜곡을 최소화하는 최적 추정자를 찾았다. 그들은 유색잡음에 오염된 신호의 상관행렬을 고유값 분해(eigen value decomposition : EVD)로 분해하기 위하여 Karhunen-Loève 변환(KLT)을 사용하였다. 신호 부공간을 나타내는 KLT 성분들은 추정자에 의해 결정되는 이득 함수에 의해 수정되는 반면, 잡음 부공간을 나타내는 나머지 KLT 성분들은 영(zero)이 된다. 이렇게 하여 개선된 신호는 수정되어진 성분들의 역 KLT에 의해 얻을 수 있다. EV는 그들의 알고리즘의 구현에서 잡음을 백색이라 가정하였다.

한편 Mittal와 Phamdo[2] 및 Rezayee 와 Gazor[3]는 Epharaim 및 Van trees의 알고리즘을 유색잡음에 적용될 수 있는 알고리즘으로 확장하였다. Mittal 및 Phamdo는 오염된 신호를 음성이 우선시되는 프레임과 잡음이 우선시되는 프레임으로 나누고 추정자를 얻기 위하여 프레임 마다 다른 KLT 행렬을 적용하였다[2]. Rezayee 및 Gazor는 유색잡음을 다루기 위해 KLT 변환된 잡음 벡터들의 상관행렬을 대각선 행렬로 근사화하는 방법으로 EV 알고리즘을 확장하였다. 이 방법에 의해 얻어진 추정자는 최적화된 추정자가 아닌 차선에 해당한다.

한편 Yi 및 Loizou(YL)는 유색잡음에 의해 오염된 음성에 대한 음성 향상을 위해 새로운 방법을 제안하였다[4]. 그들은 깨끗한 음성신호의 상관행렬과 유색잡음의 상관행렬을 동시에 대각행렬로 바꿀 수 있는 방법에 근거를 두는 비일원(nonunitary) 변환을 사용하여, 잡음에 오염된 신호를 신호-잡음(signal-plus-noise) 부공간과 잡음 부공간으로 분리하였다. 깨끗한 신호는 잡음부공간 성분을 영으로 하고 신호부공간 성분을 취하여 재구성 할 수 있었다. 이 방법은 알고리즘 자체에 유색잡음신호를 백색화시키는 과정이 포함되어 있기 때문에 가장 일반화된 알고리즘이라고 볼 수 있다. 잡음이 백색일 경우 YL 알고리즘은 EV 알고리즘이 된다.

최근에, Lee 및 Son(LS)[5] 등은 유색잡음에 오염된 신호의 음성 향상을 위해, 잡음에 오염된 음성신호의 전처리로 백색화 변환(whitening transformation : WT)을 이용하는 방법을 제안했다. 음성과 유색잡음이 서로

상관이 없다면 음성신호와와는 상관없이 유색잡음은 백색화 시킬 수 있다. 이때 음성신호는 왜곡될 수 있지만 WT에 의한 전처리 다음에 수행되는 음성 향상 후 역 WT에 의해 이 음성왜곡은 원상 복구될 수 있다. 또, WT 후에 수행되는 음성 향상은 EV접근법이나 YL접근법 중 어느 것이나 택할 수 있었다.

본 연구에서는 YL 접근법에 VAD(voice activity detector)를 구현하는 수정된, 알고리즘을 제안한다. 이렇게 수정된 알고리즘은 YL 접근법 및 LS 접근법과 컴퓨터 실험으로 성능을 비교하였다.

II. 시간영역 부공간 접근법

본 장에서는 본 연구의 기본이론을 참고문헌[1,2,4,5]로부터 요약한다.

2.1. 원리

잡음이 섞이지 않았을 때 선형 신호 모델 x 는

$$x = \Psi \cdot s \quad (1)$$

로 표현할 수 있다.

여기서 Ψ 는 랭크가 M이고 $K \times M$ 사이즈의 행렬이다. 또, $M < K$ 이고 s 는 사이즈가 $M \times 1$ 인 벡터이다. x 의 공분산 행렬(covariance matrix) R_x 는,

$$R_x = E\{x \cdot x^T\} = \Psi \cdot R_s \cdot \Psi^T \quad (2)$$

로 쓸 수 있는, 여기서 R_s 는 벡터 s 의 공분산 행렬이고 각 인자들은 양수 값을 가진다. R_x 의 랭크는 M이고, 따라서 R_x 는 K-M개의 영인 고유치(eigenvalue)을 갖는다.

깨끗한 신호 x 와 잡음신호 n 가 서로 상관관계에 없다고 하면, 오염된 신호 y 는,

$$y = \Psi \cdot s + n = x + n \quad (3)$$

가 된다. 여기서 y , x 및 n 는 각각 K-차원의 오염된 음성신호벡터, 오염 안 된 음성신호 벡터 및 잡음신호 벡

터이다. $\hat{x} = H \cdot y$ 를 깨끗한 음성신호 x 의 선형 추정자(linear estimator)라하고, H 는 $K \times K$ 행렬이다. 이 추정에 의해서 얻어지는 오차신호 ϵ 는,

$$\epsilon = \hat{x} - x = (H - I) \cdot x + H \cdot n = \epsilon_x + \epsilon_n \quad (4)$$

가 된다. 여기서 ϵ_x 는 음성왜곡을 나타내고 ϵ_n 는 잔여 잡음을 나타낸다. 신호왜곡 에너지 $\overline{\epsilon_x^2}$ 는

$$\overline{\epsilon_x^2} = E[\epsilon_x^T \epsilon_x] = \text{tr}(E[\epsilon_x \epsilon_x^T]) \quad (5)$$

이고, 잔여잡음 에너지 $\overline{\epsilon_n^2}$ 는,

$$\overline{\epsilon_n^2} = E[\epsilon_n^T \epsilon_n] = \text{tr}(E[\epsilon_n \epsilon_n^T]) \quad (6)$$

이다. 널리 알려진 시간영역 제한 최적화 문제[1, 3],

$$\begin{aligned} & \min_H \overline{\epsilon_x^2} \\ & \text{subject to: } \frac{1}{K} \overline{\epsilon_x^2} \leq \sigma^2 \end{aligned} \quad (7)$$

의 해를 구하면 최적 선형 추정자 H 를 얻을 수 있다. 여기서 σ^2 는 양의 실수이다. 식(7)의 해는

$$H_{opt} = R_x (R_x + \mu R_n)^{-1} \quad (8)$$

이다[1]. 여기서 R_x 및 R_n 은 각각 깨끗한 음성 및 잡음의 공분산 행렬이다. μ 는 라그랑지 승수(Lagrange multiplier)이다.

식(8)은 고유값 분해 $R_x = U \Delta_x U^T$ 을 이용하면

$$H_{opt} = U \Delta_x (\Delta_x + \mu U^T R_n U)^{-1} U^T \quad (9)$$

로 간략화 된다.

여기서 U 는 일원(unitary) 고유벡터 행렬이고 Δ_x 는 R_x 의 대각선 고유치 행렬이다. $R_n = \sigma_n^2 I$ 인 백색잡

음일 경우, 식(9)는 Ephraim 및 Van Tree 선형 추정자와 일치한다[1]. 행렬 $U^T R_n U$ 의 대각선 행렬 Δ_n 는,

$$\Delta_x = \text{diag}(E(|u_1^T n|), E(|u_2^T n|), \dots, E(|u_k^T n|)) \quad (10)$$

로 근사시킨다[3]. 여기서 u_k 는 R_x 의 k 번째 고유벡터이다. 그리고 음성신호의 음성결여 세그먼트로 추정된 잡음벡터이다. 식(10)의 근사로 식(9)는

$$H_{opt} \approx U \Delta_x (\Delta_x + \mu \Delta_x)^{-1} U^T \quad (11)$$

가 된다[3].

근사식 식(10)을 사용했기 때문에 유도된 추정자는 준 최적값(Suboptimal)이 된다[3]. 유색잡음에 대해서도 최적 추정자를 유도 할 수 있다[4]. 식(9)에 있는 행렬 $U^T R_n U$ 는 대각선 행렬에 가까우나, 대각선 행렬은 아니다. 행렬 U 는 대칭행렬 R_x 의 고유벡터 행렬로서 R_x 을 대각선 행렬로 만드나 R_n 을 대각선 행렬화 하지는 않는다. $U^T R_n U$ 를 식(10)에 의해 근사시키는 대신에 R_x 및 R_n 을 동시에 대각선 행렬화 시킬 수 있는 행렬을 구했다[4]. 그러한 행렬 V 는 존재하고,

$$V^T R_x V = \Lambda_x, \quad V^T R_n V = I \quad (12)$$

가 된다[6]. 여기서 Λ_x 및 V 는 각각 $\Sigma = R_n^{-1} R_x$ 즉,

$$\Sigma V = V \Lambda_x \quad (13)$$

의 고유치 행렬 및 고유벡터 행렬이다. R_n 이 양수 값을 가지면 R_x 는 실수 행렬이 된다[7]. 고유벡터 행렬 V 는 직교행렬이 아니다. R_x 의 랭크가 M 이기 때문에 행렬 Σ 도 랭크가 M 이다. 식(8)에 Σ 도 고유값 분해를 적용하고 식(12)을 이용하면 최적 추정자는

$$\begin{aligned} H_{opt} &= R_n V \Lambda_x (\Lambda_x + \mu I)^{-1} V^T \\ &= V^{-T} \Lambda_x (\Lambda_x + \mu I)^{-1} V^T \end{aligned} \quad (14)$$

로 얻어진다. 여기서, 상수 μ 는,

$$\sigma^2 = \frac{1}{K} \text{tr}\{(V^T V)^{-1} \Lambda_x^2 (\Lambda_x + \mu I)^{-2}\} \quad (15)$$

를 만족해야한다.

깨끗한 음성신호의 추정신호 \hat{x} 는 잡음이 섞인 신호 y 에 V^T 를 사용하여 변환시키고, 이득함수를 $V^T y$ 의 성분들에 곱하고, 여기에 수정된 성분들을 역변환시켜 얻을 수 있다. 그림 1은 이 과정들을 설명한다.

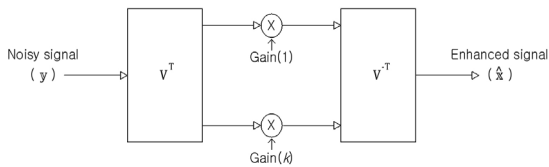


그림 1. 신호 부공간 선형 추정자
Fig. 1 Signal subspace linear estimator

이득 행렬 $G = \Lambda_x (\Lambda_x + \mu I)^{-1}$ 의 대각선 행렬이고, 이거의 k 번째 대각선 요소 g_{kk} 는,

$$g_{kk} = \begin{cases} \frac{\lambda_x^{(k)}}{\lambda_x^{(k)} + \mu}, & k = 1, 2, \dots, M \\ 0, & k = M+1, \dots, K \end{cases} \quad (16)$$

이다. 여기서 $\lambda_x^{(k)}$ 는 고유치 행렬 Λ_x 의 k 번째 대각선 요소이고, M 은 행렬 Σ 의 랭크이고 음성신호 부공간의 차원이다. 이 경우 $V^T y$ 는 y 의 KLT이 아니다. 그러나, 잡음신호가 백색잡음이면 $V^T y$ 는 y 의 KLT이 된다.

식 (14)에서 얻어진 추정자와 참고문헌 [1]에서 잡음이 백색일 때 얻어지는 선형 추정자를 비교하면 두 추정자는 같은 형태를 가진다는 것을 알 수 있다. Ephraim 및 Van Tree[1]의 추정자는 식 (4)에서 주어진 추정자의 특별한 경우가 된다. 백색잡음 ($R_n = \sigma_n^2 I$)인 경우 Σ 의 고유벡터 행렬 V 는 R_x 의 일원 고유벡터 행렬 U 가 된다. 이것은 $\Sigma = (1/\sigma_n^2)R_x$ 및 대각선 행렬 Λ_x 가 $(1/\sigma_n^2)\Delta_x$ 가 되기 때문이다. Δ_x 는 R_x 의 대각선 고

유치 행렬이다. 따라서 백색잡음인 경우 식 (14)는

$$H_{opt} = U \Delta_x (\Delta_x + \mu \sigma_n^2 I)^{-1} U^T \quad (17)$$

로 표시되고 Ephraim 및 Van Tree[1] 추정자이다. 선형 추정자 식 (14)는 백색잡음이나 유색잡음에 모두 적용될 수 있는 추정자이고, 참고문헌 [1]에 개발되었던 부공간 접근법의 일반화가 식 (14)에 주어진 추정자이다.

식 (14)의 실제 구현에는 행렬 Σ 를 추정해야 하고, 실제에서는 깨끗한 음성신호의 공분산 행렬을 구할 수 없다. 따라서, 음성신호가 잡음과 상관관계가 없다고 가정하면

$$R_y = R_x + R_n \quad (18)$$

을 얻게되고, 요구되는 Σ 는,

$$\begin{aligned} \Sigma &= R_n^{-1} R_x = R_n^{-1} (R_y - R_n) \\ &= R_n^{-1} R_y - I \end{aligned} \quad (19)$$

이다.

2.2. μ 값의 추정

식 (16)에 주어진 이득 함수의 μ 값은 실제 알고리즘의 구현을 위해 추정되어야 하고, 잔여잡음과 음성왜곡 사이에서 상충되기 때문에 추정된 음성신호 Λ_x 의 음질을 좌우한다. μ 값이 크면 잔여잡음을 많이 제거되나, 음성왜곡 또한 커진다. 역으로 μ 값을 적게 선택하면 음성왜곡은 최소화 할 수 있으나, 잔여잡음은 커진다. 따라서 적절한 μ 값을 선택해 잔여잡음과 음성왜곡 사이에서 타협점을 찾아야한다.

음성신호는 배경잡음에 대해서 마스킹 효과가 있기 때문에 음성신호가 강조된 프레임에서는 음성왜곡을 최소화해야하고, 반면 배경잡음이 강조된 프레임에서는 잔여잡음을 감소시켜야 한다. μ 값을 단구간 SNR에 의존하도록 만들어야 한다. 따라서 μ 값은,

$$\mu = \mu_0 - (SNR_{dB})/s \quad (20)$$

에 의해 추정한다. 여기서, μ_0 및 s 는 실험적으로 선택해야할 상수[4]이고, $SNR_{dB} = 10\log_{10}SNR$ 이다. 스펙트럼 차감법에 과차감 요소(over-subtraction factor)를 추정하기 위해 식 (20)이 사용되었다. 또한 프레임 SNR에 관계없이 고정된 μ 값을 사용한 경우도 있다 [1,8].

식 (12)에서, 고유치 $\lambda_x^{(k)}$ 는 고유벡터 V_k [즉, $\lambda_x^{(k)} = E(|V_k^{-T}x|^2)$]에 상응하는 신호 전력이다. 그러므로, SNR값의 추정은 변환된 영역에서,

$$SNR = \frac{tr(V^T R_x V)}{tr(V^T R_n V)} = \frac{\sum_{k=1}^M \lambda_x^{(k)}}{K} \quad (21)$$

에 의해 구할 수 있다[4].

III. 제안하는 알고리즘

제안하는 알고리즘은 YL 접근법[4] 및 LS 접근법[5]에 VAD를 구현하는 수정된 알고리즘이다. 제안하는 알고리즘은 잡음에 오염된 신호 y 에 대해 매 프레임마다 다음 일곱 단계의 과정을 거친다.

Step 1: 매 프레임의 에너지를 구한다. 즉, 식(22)에서 α 값이 임계치 α_{TH} 보다 적으면 그 프레임을 잡음 프레임으로 보고 잡음 공분산 행렬 R_n 을 구한다. 여기서 α_{TH} 는 실험으로 정해진다.

$$\alpha = \sum_{n=0}^K |y(n)|^2 \quad (22)$$

Step 2: 잡음이 섞인 음성신호의 공분산 행렬 R_y 을 구하고 식 (19)을 사용하여 행렬 Σ 을 추정한다.

Step 3: Σ 을 고유분해(EVD)한다.

$$\Sigma V = V \Lambda_x \quad (23)$$

Step 4: Σ 의 고유치가 큰 순서대로 놓았다면 (즉,

$\lambda_x^{(1)}, \lambda_x^{(2)}, \dots, \lambda_x^{(k)}$), 다음 식 (24)을 이용하여 음성신호 부공간의 차원 M 을 구한다.

$$M = \arg \max_{1 \leq k \leq K} \{\lambda_x^{(k)} > 0\} \quad (24)$$

Step 5: 다음 식 (25)에 따라 μ 값을 추정한다.

$$\mu = \begin{cases} \mu_0 - (SNR_{dB})/s, & -5 < SNR_{dB} < 20 \\ 1, & SNR_{dB} \geq 20 \\ 5, & SNR_{dB} \leq -5 \end{cases} \quad (25)$$

여기서 $\mu_0 = 4.2$, $s = 6.25$ 이다. SNR 은 식 (21)에서와 같이 구한다.

Step 6: 최적 선형 추정자를 계산한다.

$$g_{kk} = \begin{cases} \frac{\lambda_x^{(k)}}{\lambda_x^{(k)} + \mu}, & k = 1, 2, \dots, M \\ 0, & k = M+1, \dots, K \end{cases} \quad (26)$$

$$G_1 = \text{diag}\{g_{11}, \dots, g_{MM}\}$$

$$\begin{aligned} H_{opt} &= R_n V \begin{bmatrix} G_1 & 0 \\ 0 & 0 \end{bmatrix} V^T \\ &= V^{-T} \begin{bmatrix} G_1 & 0 \\ 0 & 0 \end{bmatrix} V^T \end{aligned}$$

Step 7: 잡음이 섞인 신호 y 을 식 (26)에 적용하여, 잡음이 제거된 음성신호를 추정한다.

IV. 컴퓨터 실험 및 성능 평가

알고리즘의 구현에는 잡음에 오염된 신호의 공분산 행렬 R_y 와 잡음신호의 공분산 행렬 R_n 을 정확하게 추정해야한다. 본 연구에서는 비편향 자기상관 시퀀스 (autocorrelation sequence)을 관측되는 신호로부터 얻었다. 이 시퀀스의 K샘플을 이용하여 Toeplitz 공분산 행렬을 구성하였다. 공분산 행렬을 구성할 때 과거나 미

래 프레임 사용하지 않았다. 음성신호와 잡음신호는 모두 16KHz에서 샘플링 되었다. K=40샘플이었다. 참고문헌 [4]에서는 잡음신호의 공분산 행렬 R_n 을 추정하기 위해 테스트 문장에 있는 묵음 구간인 초기 몇 프레임을 사용하여 R_n 을 추정하였으며, VAD를 사용하지 않았다. 그러나, 본 논문에서는 제 3장 2절의 단계 1 및 2에서 언급한 바와 같이 프레임마다 그 프레임의 에너지를 식 (26)과 같이 구하여 실험적으로 정해지는 임계값보다 작은 값이면 그 프레임을 잡음 프레임으로 간주하여 R_n 을 갱신하였다.

알고리즘의 구현에는 공분산 행렬을 추정하기 위하여 사각 윈도우를 사용하였고, 매 프레임을 50% 중첩시켰다. 잡음에 오염된 음성신호는 최종적으로 해밍(Hamming) 윈도우와 음성 향상된 후 중첩 및 가산(overlap-and-add) 접근법을 사용하여 복구하였다.

4.1. 평가척도

성능평가는 음성 스펙트럼 왜곡(speech spectral distortion : SSD)과 음성 향상 시스템 출력에서의 신호 대 잡음비인 SNR을 평가 척도로 이용하였다. 먼저 2차원 SSD 척도는 다음과 같이 정의된다[2]. a 및 b 를 N 차원 벡터화 한다. 각 벡터는 먼저 에너지가 1(0dB)이 되도록 정규화 한다. 그리고 -30dB의 에너지를 가지는 또 다른 백색잡음 벡터 c 를 벡터 a 및 b 에 보탠다. 이 백색 잡음 벡터 c 를 보태는 이유는 내수적으로 계산될 척도에서 $\log(0)$ 의 계산을 막기 위함이다. 따라서 벡터 a 및 b 는 각각 $\tilde{a} = a / \|a\| + c$ 및 $\tilde{b} = b / \|b\| + c$ 가 된다. 벡터 \tilde{a} 및 \tilde{b} 는 길이가 64인 중첩되지 않는 프레임들로 나눈다. 나누어진 각 64샘플 프레임에 영을 192개 보태어 영 첨가(zero-padding) 한다. 이렇게 하여 생긴 256샘플의 매 프레임의 DFT(discrete Fourier transform)를 계산한다. $\tilde{A}_p(k)$ 및 $\tilde{B}_p(k)$ 를 각각 벡터 \tilde{a} 및 \tilde{b} 의 DFT의 P번째 프레임의 k번째 주파수 성분이라 한다. 벡터 a 및 b 간의 SSD는,

$$s(a, b) = \frac{1}{p} \frac{1}{256} \sum_{i=1}^p \times \sum_{k=0}^{255} 20|\log|\tilde{A}_p(k)| - \log|\tilde{B}_p(k)|| \quad (27)$$

로 정의 된다. 여기서 p 는 관측된 음성신호의 총 프레임 수이다. 제안하는 추정자는 선형이기 때문에 식 (3) 및 식 (14)로부터,

$$\tilde{y} = H_y = H_x + H_n \quad (28)$$

로 나누어 질 수 있다. 여기서 H_x 는 신호 부분이고 H_n 은 잡음 부분이다. 이상적으로 볼 때 $s(x, H_x) = 0$ 이 되도록 하는 선형 추정자 H 가 필요하다. 이 $s(x, H_x)$ 을 음성 스펙트럼 왜곡이라 정의한다. 잡음에 오염된 신호 y 의 신호대잡음비(SNR)은,

$$SNR_{dB} = 10\log_{10} \frac{\sum_{t=1}^N \sum_{k=1}^K x_t^2(k)}{\sum_{t=1}^N \sum_{k=1}^K (y_t(k) - x_t(k))^2} \quad (29)$$

로 정의 되었고, 여기서 N은 관측된 신호 y 의 총 프레임 수이고 $y_t(k)$ 는 t 번째 프레임의 k번째 샘플이다. 음성 향상 시스템의 출력 SNR은 식 (28)로부터 H_x 및 H_y 을 사용하여 식 (29)로 계산되었다.

4.2. 실험결과 및 성능평가

성능평가를 위해 5명의 남자와 5명의 여자가 발음한 문장 “She had your dark suit in greasy wash water all year”을 TIMIT 데이터베이스로부터 발췌하여 입력 음성 데이터로 사용하였다. 음성 데이터의 시작 부분에서 0.05초 미만 구간의 큰 잡음이 있어 전처리로 제거한 후 사용하였으며, 샘플링율은 16 kHz이다. 깨끗한 음성에 자동차 잡음과 다중화자 배블(multi-talker babble) 잡음은 AURORA 데이터베이스로부터 취하여 SNR이 5dB이 되도록 음성 데이터에 각각 더하여 입력으로 사용하였다. 음성 출력 신호의 SNR은 식 (29)를 사용하여 계산하였고, 음성 스펙트럼 왜곡은 식 (27)을 사용하여 계산하였다.

전체실험은 남성 화자 5명과 여성화자 5명에 대해 각각 실험하고 평균한 결과를 표 1에서 표 4까지 나타내었다.

표 1 및 표 2는 깨끗한 음성에 자동차 잡음을 SNR이 5dB이 되도록 첨가한 경우의 실험결과이다. 실험결과에서 남성 화자 항은 5명의 남자가 발음한 문장을 실험하여 평균하였다. 같은 방법으로 여성 화자 항은 여자 5명이 같은 문장을 발음한 것을 실험하여 평균하였다. 표 1에서 볼 수 있듯이 자동차 잡음일 경우 출력 SNR이 VAD가 적용된 제안한 알고리즘이 기존의 알고리즘[5]보다 남성음성의 경우 약 0.86dB, 여성음성일 경우 약 0.64dB 정도 향상되었다. 그리고, 표 2에서 볼 수 있듯이 음성 스펙트럼 왜곡의 경우 제안하는 알고리즘이 기존 알고리즘[5]보다 남성의 경우 0.08, 여성의 경우도 0.08 낮았다.

표 3 및 표 4는 깨끗한 음성에 다중화자 배블을 SNR이 5dB이 되게 첨가하였을 경우의 실험결과이다. 실험 방법은 자동차 잡음일 경우와 동일하게 실행하였다. 표 3에 언급한 바와 같이 다중화자 배블인 경우 입력 SNR이 5dB일 때 기존 알고리즘[5]보다 VAD가 적용된 제안한 알고리즘의 출력 SNR이 남자의 경우 약 0.14dB, 여자의 경우는 약 0.04dB 향상되었다. 그러나 표 4에서 보는 바와 같이 다중화자 배블의 경우 음성 스펙트럼 왜곡(SSD)은 기존 알고리즘[5]보다 남성의 경우 0.03, 여성의 경우 0.01 향상되었다.

표 1. 5dB 자동차 소음 첨가한 경우의 성능 비교(전체 평균 SNR)
Table. 1 Comparison for adding 5dB car noise(total average SNR)

알고리즘 종류	남성 화자	여성 화자
YL 알고리즘[4]	7.67	7.32
LS 알고리즘[5]	8.42	8.33
LS 알고리즘 +VAD[5]	9.34	9.29
제안하는 알고리즘	10.2	9.93

표 2. 5dB 자동차 소음 첨가한 경우의 성능 비교(평균 SSD)
Table. 2 Comparison for adding 5dB car noise(average SSD)

알고리즘 종류	남성 화자	여성 화자
YL 알고리즘[4]	0.68	0.65
LS 알고리즘[5]	0.74	0.68
LS 알고리즘 +VAD[5]	0.68	0.65
제안하는 알고리즘	0.60	0.57

표 3. 5dB 다중화자 배블 잡음 첨가한 경우의 성능 비교(전체 평균 SNR)

Table. 3 Comparison for adding 5dB multi-talker babble(total average SNR)

알고리즘 종류	남성 화자	여성 화자
YL 알고리즘[4]	8.28	8.06
LS 알고리즘[5]	7.67	7.33
LS 알고리즘 +VAD[5]	8.37	8.04
제안하는 알고리즘	8.51	8.08

표 4. 5dB 다중화자 배블 잡음 첨가한 경우의 성능 비교(평균 SSD)

Table. 4 Comparison for adding 5dB multi-talker babble(average SSD)

알고리즘 종류	남성 화자	여성 화자
YL 알고리즘[4]	0.62	0.58
LS 알고리즘[5]	0.69	0.66
LS 알고리즘 +VAD[5]	0.67	0.62
제안하는 알고리즘	0.64	0.61

V. 결 론

본 논문에서는 유색잡음에 오염된 음성신호의 음성 향상을 위해, YL 알고리즘에 VAD를 구현하는 수정된 알고리즘을 제안했다. 최근 개발된 기존 알고리즘[5]과 컴퓨터 시뮬레이션으로 성능을 비교검토하였다.

- 1) 유색잡음이 자동차 잡음인 경우 (입력 SNR은 5dB)
 - ① 출력 SNR은 남성음성의 경우 기존 알고리즘보다 제안한 알고리즘이 0.86dB 향상되었고, 여성 음성의 경우는 기존 알고리즘보다 제안한 알고리즘이 0.64dB 향상되었다.
 - ② 음성 스펙트럼 왜곡(SSD)은 남성음성과 여성음성의 경우 기존 알고리즘이나 제안한 두 알고리즘이 각각 0.08 향상되었다.
- 2) 유색잡음이 다중화자 배블 잡음인 경우 (입력 SNR은 5dB)
 - ① 출력 SNR은 기존 알고리즘과 제안한 알고리즘은 남성과 여성음성에 대해서 각각 0.14dB 및 0.04dB 정도 향상되었다.

- ② 음성 스펙트럼 왜곡(SSD)은 기존 알고리즘 보다 제안한 알고리즘이 남성 및 여성의 경우 각각 0.03 및 0.01 향상되었다.

감사의 글

“이 논문은 부산대학교 자유과제 학술연구비 (2년)에 의하여 연구되었음.”

REFERENCES

[1] Y. Ephraim and H. L. Van Trees, “A signal subspace approach for speech enhancement,” *IEEE Trans. Speech Audio Processing*, vol. 3, pp.251 - 266, 1995.

[2] U. Mittal and N. Phamdo, “Signal/noise KLT based approach for enhancing speech degraded by colored noise,” *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 159 - 167, Mar. 2000.

[3] A. Rezayee and S. Gazor, “An adaptive KLT approach for speech enhancement,” *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 87 - 95, Feb. 2001.

[4] Yi Hu and Philipos C. Loizou, A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise *IEEE Trans. Speech Audio Processing*, vol. 11, no. 4, pp. 334-341. July 2003.

[5] J. W. Lee, K. S. Son, J. S. Park, H. T. Kim, “A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise Using Whitening Transformation”, *The Journal of the Korean Institute of Information and Communication Engineering*, vol. 15, no. 8, pp. 1665-1674, 2011.

[6] S. B. Searle, *Matrix Algebra Useful for Statistics*. New York: Wiley, 1982.

[7] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. New York: Harcourt Brace Jovanovich, 1988.

[8] Y. Hu and P. C. Loizou, “A subspace approach for enhancing speech corrupted by colored noise,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, Orlando, FL, May 2002, pp. 573 - 576.



손경식(Kyung-sik Son)

1973년 2월 부산대학교 전자공학과(학사)
1977년 8월 부산대학교 전자공학과(석사)
1991년 8월 경북대학교 전자공학과(박사)
1986년 10월~현재 부산대학교 전자공학과 교수
※관심분야 : 적응신호처리, 음향신호처리



김현태(Hyun-Tae Kim)

1989년 2월 부산대학교 전자공학과(학사)
1995년 2월 부산대학교 전자공학과(석사)
2000년 2월 부산대학교 전자공학과(박사)
2002년 3월~현재 동의대학교 멀티미디어공학과 교수
※관심분야 : 적응신호처리, 음향신호처리, 멀티미디어신호처리