

변수내장형 다채널 위너필터를 위한 목적신호대잡음 기여비를 이용한 잡음추정기법

홍정표*

Noise Statistics Estimation Using Target-to-Noise Contribution Ratio for Parameterized Multichannel Wiener Filter

Jungpyo Hong*

*Assistant professor, Department of Information and Communication Engineering, Changwon National University, Changwon, 51140 Korea

요 약

변수내장형 다채널 위너 필터는 내장된 변수를 이용하여 잔여잡음과 신호왜곡 간의 트레이드오프를 조절할 수 있는 선형 필터이다. 이러한 변수내장형 다채널 위너필터를 적용하기 위해서는 정확한 잡음추정이 중요하데 널리 쓰이는 다채널 최소 제어 재귀 평균 기법이 있다. 하지만 다채널 최소 제어 재귀 평균 기법은 방향성 간섭 신호가 존재할 경우 잡음추정의 정확도가 하락하여 변수내장형 다채널 위너필터의 성능이 저하되는 문제점이 있다. 따라서, 본 논문에서는 변수내장형 다채널 위너필터를 위한 새로운 잡음 추정 기법을 제안한다. 제안한 방법은 주로 잡음 섞인 마이크로폰 입력 신호의 전력 스펙트럼 밀도에 대해 고유값 분해, 방향성 정보를 이용한 목적신호의 기여도 추정, 목적신호의 기여도를 보다 합리적으로 추정하기 위한 지수 가중치 부가의 일련의 과정을 수행한다. 제안한 방법을 평가하기 위해 신호대잡음비, 음성왜곡도 등의 총 4가지 객관적 성능 평가 방법을 이용하여 기존의 방법과 비교하였다. 실험을 통해 방향성 간섭신호가 존재하는 환경에서 제안한 잡음 추정기법을 적용한 다채널 위너필터의 성능이 향상됨을 확인하였다.

ABSTRACT

Parameterized multichannel Wiener filter (PMWF) is a linear filter that can control the trade-off between residual noise and signal distortion using the embedded parameter. To apply the PMWF to noisy inputs, accurate noise estimation is important and multichannel minima-controlled recursive averaging (MMCRA) is widely used. However, in the case of the MMCRA, the accuracy of noise estimation decreases when a directional interference is involved into the array inputs. Consequently, the performance of the PMWF is degraded. Therefore, we propose a noise power spectral density (PSD) estimation method for the PMWF in this paper. The proposed method is based on a consecutive process of eigenvalue decomposition on noisy input PSD, estimation of the target component contribution using directional information, and exponential weighting for improved estimation of the target contribution. For evaluation, four objective measures were compared with the MMCRA and we verify that the PMWF with the proposed noise estimation method can improve performance in environments where directional interferences exist.

키워드 : 마이크로폰 배열, 잡음추정, 변수내장형 다채널 위너필터, 음성강화

Keywords : Microphone array, Noise estimation, Parameterized multichannel Wiener filter, Speech Enhancement

Received 1 December 2022, Revised 7 December 2022, Accepted 12 December 2022

* Corresponding Author Jungpyo Hong(E-mail: hansin@changwon.ac.kr, Tel:+82-55-213-3838)

Assistant Professor, Department of Information and Communication Engineering, Changwon National University, Changwon, 51140 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.12.1926>

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. 서론

음성인식시스템의 발달로 스마트 TV, 냉장고, 세탁기, 청소기 등의 생활 가전과 로봇, 자동차등 음성인터페이스의 사용이 대중화되고 있다. 하지만 이런 다양한 환경에서 음성인터페이스를 사용할 경우 음성과 함께 필연적으로 잡음이 마이크로폰 입력으로 개입하게 된다. 이를 효과적으로 제거하여 음성의 질을 향상시키기 위한 음성강화(speech enhancement) 연구는 오래전부터 현재까지 활발히 진행되고 있다 [1-11]. 다양한 연구 중에서 빔포밍과 같이 다채널 마이크로폰을 사용하여 방향성 간섭신호를 효과적으로 제거하기 위한 연구가 한가지 주류를 이루고 있다 [1-5]. 특히, 2채널 잡음제거 기법의 경우 스마트폰, 보청기와 같이 장치의 크기가 작고, 마이크로폰 수를 최소한으로 절감하면서 마이크로폰 간의 시간지연 정보를 활용할 수 있다는 점에서 활발히 연구되고 있다 [4-5]. 한편, 최근에 딥러닝 기술의 발달로 음성강화 분야에도 적용되면서 딥러닝 기반의 다채널 잡음제거 기술이 연구되고 있다 [6-10]. 단채널 마이크로폰을 사용한 기법임에도 불구하고 성능이 매우 뛰어나다는 장점이 있지만 딥러닝 모델을 훈련하기 위해 많은 데이터가 필요하고 훈련시간, 비용 등이 많이 든다는 점에서 단점이 있다.

변수내장형 다채널 위너필터(Parameterized Multichannel Wiener Filter, PMWF)는 선형 필터로서 마이크로폰 배열을 이용한 음성강화를 위한 최신방법 중 하나이다 [11]. 변수내장형 다채널 위너필터는 주파수영역 잡음제거 기법으로 잔여 잡음과 신호 왜곡의 트레이드오프(trade-off)를 조절할 수 있다. 변수내장형 다채널 위너필터는 전력 스펙트럼 밀도(Power Spectral Density, PSD)와 같은 신호와 잡음의 통계치를 이용하여 계산할 수 있고 널리 알려진 최소 분산 무왜곡 응답(Minimum Variance Distortionless Response, MVDR) 필터는 변수내장형 다채널 위너필터의 특수한 형태이다.

다채널 음성존재확률(Multichannel Speech Presence Probability, MSPP) 기반의 잡음 추정 방법인 다채널 최소 제어 재귀 평균(Multichannel Minima-Controlled Recursive Averaging) 기법 [12]은 단채널 최소 제어 재귀 평균 기법 [13]의 다채널 확장 형태로 주변소음(ambient noise)을 추정 및 추적하는 성능이 뛰어나고 변수내장형 다채널 위너필터의 잡음제거 기법으로 많이

사용된다. 하지만 다채널 최소 제어 재귀 평균기법은 목적신호와 유사한 특성을 갖는 방향성 간섭 신호가 존재할 경우 잡음 추정 성능이 하락한다. 그 결과 부정확한 잡음추정으로 인해 변수 내장형 다채널 위너필터의 성능이 저하된다.

따라서, 본 논문에서는 마이크로폰 배열 응답 벡터(array response vector)를 이용한 새로운 방식의 잡음추정기법을 제안한다. 제안한 방법은 주로 잡음 섞인 마이크로폰 입력 신호의 전력 스펙트럼 밀도에 대해 고유값 분해, 방향성 정보를 이용한 목적신호의 기여도 추정, 목적신호의 기여도를 보다 합리적으로 추정하기 위한 지수 가중치 부가의 일련의 과정을 수행한다. 제안한 잡음 추정 기법을 이용하여 변수내장형 다채널 위너필터를 수행할 경우 방향성 간섭신호가 존재하는 환경에서도 성공적으로 잡음을 제거할 수 있다.

본 논문은 2장에서 관련 연구 및 신호 모델을 제시하고 3장에서 제안한 잡음 추정 기법을 자세히 설명하며 4장에서 몇가지 실험을 통해 제안한 방법의 유효성을 입증하고 5장에서 결론을 제시한다.

II. 신호 모델 및 관련 연구

2.1. 신호 모델

본 논문에서는 m 번째 잡음 섞인 마이크로폰 입력의 단시간푸리에변환(Short Time Fourier Transform, STFT)을 $Y_m(t, k)$ 로 정의하면

$$Y_m(t, k) = X_m(t, k) + N_m(t, k), \quad m = 1, 2, \dots, M \quad (1)$$

와 같이 나타낼 수 있다. 여기서 $X_m(t, k)$ 과 $N_m(t, k)$ 은 각각 방향성 신호와 주변잡음(ambient noise)을 의미한다. 또한, t, k, M 은 각각 시간, 주파수 인덱스와 마이크로폰 개수를 말한다. 위 수식 (1)을 벡터-행렬 표기법으로 나타내면

$$\begin{aligned} \mathbf{y}(t, k) &= \mathbf{x}(t, k) + \mathbf{n}(t, k) \\ &= \mathbf{H}(k)\mathbf{s}(t, k) + \mathbf{n}(t, k) \end{aligned} \quad (2)$$

와 같다. 여기서 $\mathbf{y}(t, k) = [Y_1(t, k), Y_2(t, k), \dots, Y_M(t, k)]^T$ 이고, $\mathbf{s}(t, k) = [S_1(t, k), S_2(t, k), \dots, S_D(t, k)]^T$ 이며, D 와 T 는 방향성 신호의 개수와 전치(transpose)를

의미한다. 또한, $\mathbf{x}(t,k)$ 와 $\mathbf{n}(t,k)$ 은 $\mathbf{y}(t,k)$ 와 유사한 형태를 갖는다. 그리고 $\mathbf{H}(k)$ 는 방향성 신호보다 훨씬 천천히 변한다는 가정하에 정의된 $M \times D$ 차원의 신호전달 행렬(signal transfer matrix)를 뜻한다.

추가적으로 잡음섞인 마이크로폰 입력의 전력스펙트럼 밀도를 정의하면

$$\Phi_{\mathbf{y}\mathbf{y}}(t,k) = E[\mathbf{y}(t,k)\mathbf{y}(t,k)^H] \quad (3)$$

와 같다. 신호와 잡음의 전력스펙트럼밀도 또한 수식(3)와 유사하게 정의 된다. 여기서 H 와 E 는 각각 헤르미산 전치 (Hermitian transpose)와 기댓값 (expectation)을 의미한다. 실제로 수식 (3)의 기댓값은 다음과 같이 재귀 평균 (recursive averaging) 방법으로 근사하여 구한다.

$$\hat{\Phi}_{\mathbf{y}\mathbf{y}}(t,k) = \alpha_{\mathbf{y}}(t,k)\hat{\Phi}_{\mathbf{y}\mathbf{y}}(t-1,k) + (1-\alpha_{\mathbf{y}}(t,k))\mathbf{y}(t,k)\mathbf{y}(t,k)^H \quad (4)$$

여기서 $\alpha_{\mathbf{y}}(t,k)$ 는 시변형(time-varying) 평활화 변수 (smoothing parameter) 이고 $\hat{}$ 은 추정치를 의미한다.

2.2. 변수내장형 다채널 위너 필터

변수내장형 다채널 위너필터는 잔여잡음 (residual noise)의 양과 신호왜곡 (signal distortion)의 정도를 조절할 수 있는 최적 선형 필터이다. 이 필터는 잔여 잡음의 양을 어느 정도 이하로 유지하면서 신호 왜곡을 최소화 시키거나 반대로 신호 왜곡을 어느 정도 이하로 유지하면서 잔여잡음의 양을 최소화 시키는 제약이 있는 최적화 (constrained optimization)를 통해 도출된다 [11]. 라그랑주 승수법 (Lagrange multiplier)을 통해 제한 최적화 문제를 풀면 변수 내장형 위너 필터는

$$\mathbf{g}^H(t,k) = [\Phi_{\mathbf{x}\mathbf{x}}(t,k) + \mu\Phi_{\mathbf{nn}}(t,k)]^{-1}\Phi_{\mathbf{x}\mathbf{z}}(t,k)\mathbf{u}_{m_0} \quad (5)$$

와 같이 유도된다. 여기서 $\mathbf{u}_{m_0} = [0,0,\dots,1,\dots,0]^T$ 는 m_0 번째에만 1인 채널 선택 벡터이다. 그리고 μ 는 신호 왜곡과 잔여잡음을 조절하는 변수로 μ 값이 크면 잔여잡음이 적지만 신호왜곡이 크고 μ 값이 작으면 반대가 된다 [11].

2.3. 다채널 최소 제어 재귀 평균

다채널 최소 제어 재귀 평균 방법은 다채널 음성존재 확률 기반의 잡음 추정 기법으로 Cohen의 단채널 최소 제어 재귀 평균 기법 [13]을 다채널로 확장한 방법이다. 최소값 통계치 (minimum statistics) 기반의 잡음추정 기법은 잡음 추정치에서 급격한 변화가 있는 환경에서도 안정적인 특징이 있다. 하지만 다채널 최소 제어 재귀 평균 기법은 단채널 잡음 추정 기법으로부터 유래된 기법으로 방향성 정보를 사용하지 않는다. 따라서, 신호와 유사한 방향성 간섭신호가 있을 경우 잡음추정치가 부정확할 수 있고 그로인해 변수내장형 다채널 위너필터의 성능이 저하될 수 있는 단점이 있다.

III. 제안한 잡음 추정 기법

3.1. 고유값 분해를 통한 목적신호 기여도 추정

잡음섞인 마이크로폰 입력의 전력 스펙트럼 밀도 행렬로부터 목적신호 성분을 추정하기 위해서, 스펙트럼 이론 (spectral theorem)에 의해 잡음섞인 마이크로폰 입력의 전력스펙트럼 밀도 행렬을 다음과 같이 직교하는 고유벡터들로 분해할 수 있다 [14].

$$\Phi_{\mathbf{y}\mathbf{y}}(t,k) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H = \sum_{m=1}^M \lambda_m \mathbf{v}_m \mathbf{v}_m^H \quad (6)$$

여기서 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M]$ 와 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ 는 각각 고유벡터 행렬과 고유값 행렬이다. 고유벡터 행렬과 고유값 행렬은 시간 t 와 주파수 k 에 영향을 받는 값이지만 수식의 간결성을 위해 이후에 나오는 수식에서는 생략하기로 한다.

잡음섞인 마이크로폰 배열 입력의 전력스펙트럼 밀도 행렬로부터 분해된 M 개의 고유값은 각각 해당하는 고유벡터 축에 대하여 목적신호 전력 성분과 잡음 전력 성분의 합으로 구성된다. 따라서, 각각의 고유벡터 축에서 목적신호 전력 성분을 추정하기 위해서 목적신호성분의 기여도 (contribution)를 정의하면

$$c_m(t,k) = |\mathbf{v}_m^H \mathbf{d}(\theta,k)| / M, \quad m = 1, 2, \dots, M \quad (7)$$

와 같다. 여기서 $\mathbf{d}(\theta,k) = [1, e^{-j\omega_k \tau_2(\theta)}, \dots, e^{-j\omega_k \tau_M(\theta)}]^T$ 는

목적신호방향의 배열응답벡터이다. 또한, $\theta, \omega_k, \tau_m(\theta)$ 는 각각 목적신호의 도래각(direction-of-arrival, DOA), k 번째 각 주파수(angular frequency), m 번째 마이크론의 시간지연(time delay)를 뜻한다. 그리고 수식 (7)은 정규화(normalization)되어 $c_m(t, k) \in [0, 1]$ 이다. 참고문헌 [15]의 기여도(contribution)는 잡음이 우세한 부공간(noise-dominant subspace)을 찾고 0 또는 1의 이진 가중치(binary weight)를 할당하여 제거(hard decision)하기 위해 사용된 반면에 수식(7)의 기여도는 잡음 섞인 입력으로부터 목적신호의 기여도와 잡음의 비율을 추정함으로써 각각의 부공간에서 목적신호 전력성분과 잡음의 전력 성분을 분리(softly split)한다. 수식 (7)을 이용하여 목적신호 성분과 그 외 잡음신호를 분리하는 과정은

$$\begin{aligned} \hat{\Phi}_{yy}(t, k) &= \hat{\Phi}_{xx}(t, k) + \hat{\Phi}_{nn}(t, k) \\ &= \hat{\Phi}_{tt}(t, k) + \hat{\Phi}_{ii}(t, k) + \hat{\Phi}_{nn}(t, k) \\ &= \hat{\Phi}_{xx}^-(t, k) + \hat{\Phi}_{nn}^-(t, k) \end{aligned} \quad (8)$$

와 같다. 여기서 $\hat{\Phi}_{xx}^-(t, k) = \hat{\Phi}_{tt}(t, k)$ 는 목적신호의 전력스펙트럼 밀도 행렬이고 $\hat{\Phi}_{nn}^-(t, k) = \hat{\Phi}_{ii}(t, k) + \hat{\Phi}_{nn}(t, k)$ 는 방향성 간섭신호와 주변잡음이 결합된 잡음 전력스펙트럼 밀도이다.

목적신호 기여도를 이용하여 재정의된 신호와 잡음 전력스펙트럼 밀도를 추정하면

$$\hat{\Phi}_{xx}^-(t, k) = \sum_{m=1}^M c_m(t, k) \lambda_m \mathbf{v}_m \mathbf{v}_m^H \quad (9)$$

$$\hat{\Phi}_{nn}^-(t, k) = \sum_{m=1}^M (1 - c_m(t, k)) \lambda_m \mathbf{v}_m \mathbf{v}_m^H \quad (10)$$

와 같다. 하지만 수식 (9)와 수식 (10)으로 추정된 전력스펙트럼 밀도는 추정된 목적신호 기여도의 주파수 의존적 차이 때문에 변수내장형 다채널 위너필터에 적용하면 저주파영역에서 불충분한 잡음제거와 고주파영역에서 과도한 신호 왜곡이 발생한다.

3.2. 지수 감소 함수를 이용한 가중치 할당

공간 응답(spatial response)의 확률 밀도 함수(Probability Density Function, PDF)는

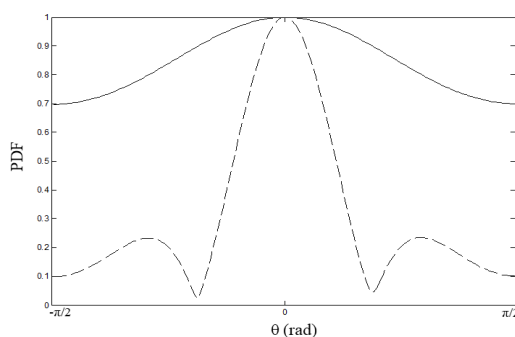


Fig. 1 Probability density function of spatial responses. (m=1, solid: 1 kHz, dashed: 3 kHz)

$$p_m(\theta, k) = \frac{c_m(t, k; \theta)}{\int_{-\pi/2}^{\pi/2} c_m(t, k; \theta) d\theta}, \quad m = 1, 2, \dots, M \quad (11)$$

와 같이 정의된다 [16]. 공간 응답의 확률밀도 함수는 그림 1과 같이 저주파에서는 정점(peak)이 넓게 분포하고 고주파에서는 좁게 분포하는 주파수에 크게 의존적인 특성이 있다. 이러한 차이를 완화시키기 위해 추정된 목적신호 기여도에 주파수별 적절한 가중치를 부가하는 것이 필요하다. 가중치를 목적신호기여도에 부가하는 것의 물리적인 의미는 공간 응답 확률밀도함수의 분포를 성형(shaping)하는 것이다. 1보다 큰 가중치를 적용하는 것은 확률 밀도함수를 뾰족하게 만드는 것이고 1보다 작은 가중치를 적용하는 것은 확률 밀도 함수를 넓게 만든다.

본 논문에서는 위 목적을 달성하기 위해 지수 감소 함수(exponentially decaying function)를 제안한다. 제안하는 지수 감소 함수는

$$w(k) = \gamma e^{-\nu k} \quad (12)$$

와 같이 정의된다. 여기서 γ 와 ν 는 각각 직류 성분을 뜻하는 초기치와 감소율(decay rate)을 나타낸다. 따라서, 가중화된 목적신호 기여도는

$$c_m^{w(k)}(t, k) = (c_{m(t, k)})^{w(k)}, \quad m = 1, 2, \dots, M \quad (13)$$

와 같이 표현된다. 수식 (13)을 수식 (9), 수식 (10)에 적용하여 추정된 신호 및 잡음 전력 스펙트럼 밀도를 이용하여 수식 (5)의 변수내장형 다채널 위너필터를 구하면 성공적인 잡음제거가 가능하다. 또한, 수식 (9), 수식 (10), 수식 (13)을 수식(5)에 적용하여 정리하면

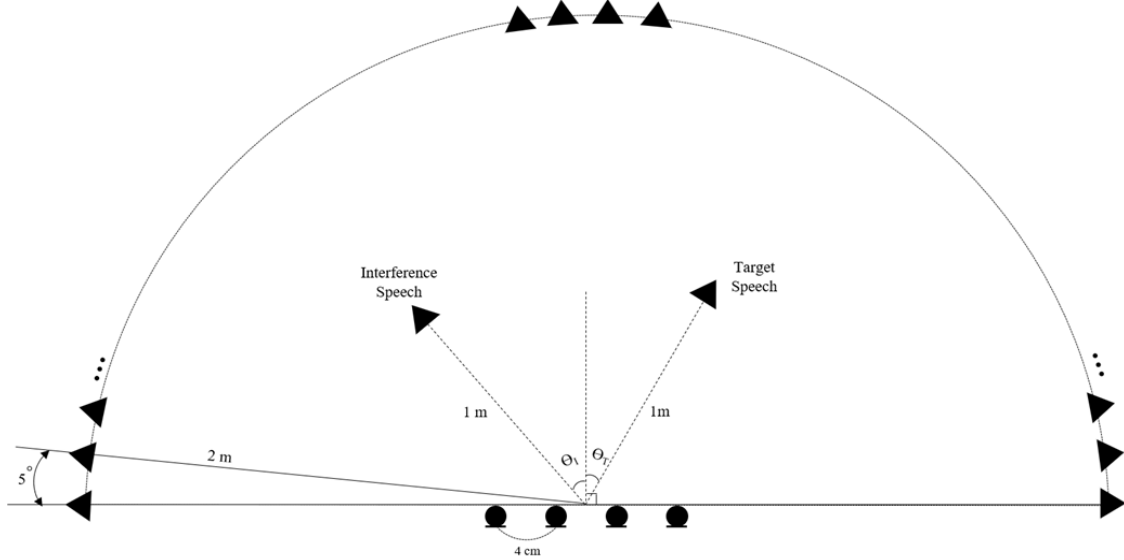


Fig. 2 Experimental environments (●: microphone, ▲: speaker)

$$\tilde{\mathbf{g}}^H(t, k) = \mathbf{V}\mathbf{\Lambda}_g \mathbf{V}^H \mathbf{u}_{m_0} \quad (14)$$

와 같이 부공간 영역에서 재해석 될 수 있다. 여기서

$$\mathbf{\Lambda}_g = \mathbf{\Lambda}_x^{(w)} \left(\mathbf{\Lambda}_x^{(w)} + \mu \mathbf{\Lambda}_n^{(w)} \right)^{-1}, \quad (15)$$

$$\mathbf{\Lambda}_x^{(w)} = \text{diag}(c_1^{w(k)}, c_2^{w(k)}, \dots, c_M^{w(k)}), \quad (16)$$

$$\mathbf{\Lambda}_n^{(w)} = \text{diag}(1 - c_1^{w(k)}, 1 - c_2^{w(k)}, \dots, 1 - c_M^{w(k)}). \quad (17)$$

수식 (14), 수식(15), 수식(16), 수식 (17)을 통해 제안한 잡음 추정 방법으로 구한 변수내장형 위너필터는 잡음 섞인 입력의 고유값과 독립적으로 가중화된 목적신호 대잡음 기여도 비율 (target-to-noise contribution ratio),

$$\rho_m = \frac{c_m^{w(k)}}{(1 - c_m^{w(k)})} \quad (18)$$

만으로 표현될 수 있음을 알게 되었다.

IV. 실험 환경 및 결과

4.1. 실험환경

그림 2는 전반적인 실험 조건을 보여준다. 마이크론 배열은 등간격 선형으로 배치되었고 마이크 간의 간

격은 0.04 m, 마이크로폰 개수는 4개였다. 데이터베이스 구축을 위해 목적신호와 방향성 간섭신호는 IEEE sentence [17] 30문장을 사용하였고 주변 잡음으로 AURORA DB [18]의 레스토랑 잡음을 사용하였다. 목적신호와 방향성 간섭신호는 같지 않게 IEEE sentence 에서 랜덤하게 선택되었다. 목적신호대 주변잡음비(Signal-to-ambient Noise Ratio, SNR)는 15 dB로 고정하고 목적신호대 방향성 간섭신호비 (Signal-to-Interference Ratio, SIR)는 각각 0, 5, 10, 15, 20 dB로 SIR별로 각각 30개의 잡음섞인 문장을 생성하였다. 각각의 음원과 마이크로폰 사이의 임펄스 응답 (impulse response)은 이미지 방법 (image method) [19]으로 모사하였고 목적신호와 방향성 간섭신호를 해당하는 임펄스 응답과 합성곱(convolution)을 통해 마이크로폰 입력을 모사하였다. 주변잡음 생성을 위해 스피커를 방향성 간섭 신호보다 더 멀리 5도 단위로 37개를 배치하였다. 37개의 스피커로부터 각각의 마이크로폰으로 입사된 주변잡음을 평균하여 주변소음을 생성하였다 [20]. 스피커의 개수가 많을수록 실제 무방향성 주변잡음과 유사하게 생성할 수 있다. 모사된 방의 크기는 5 m × 4 m × 3 m 이고 반향 시간 (reverberation time)은 300 ms 이었다. 수집된 잡음 섞인 마이크로폰 입력 데이터베이스는 8 kHz 표본화율과 16비트 해상도이다. 또한, 신호처리를 위한 프레임 크기와 고속 푸리에 변환 크기는 256이었고, 목적신호

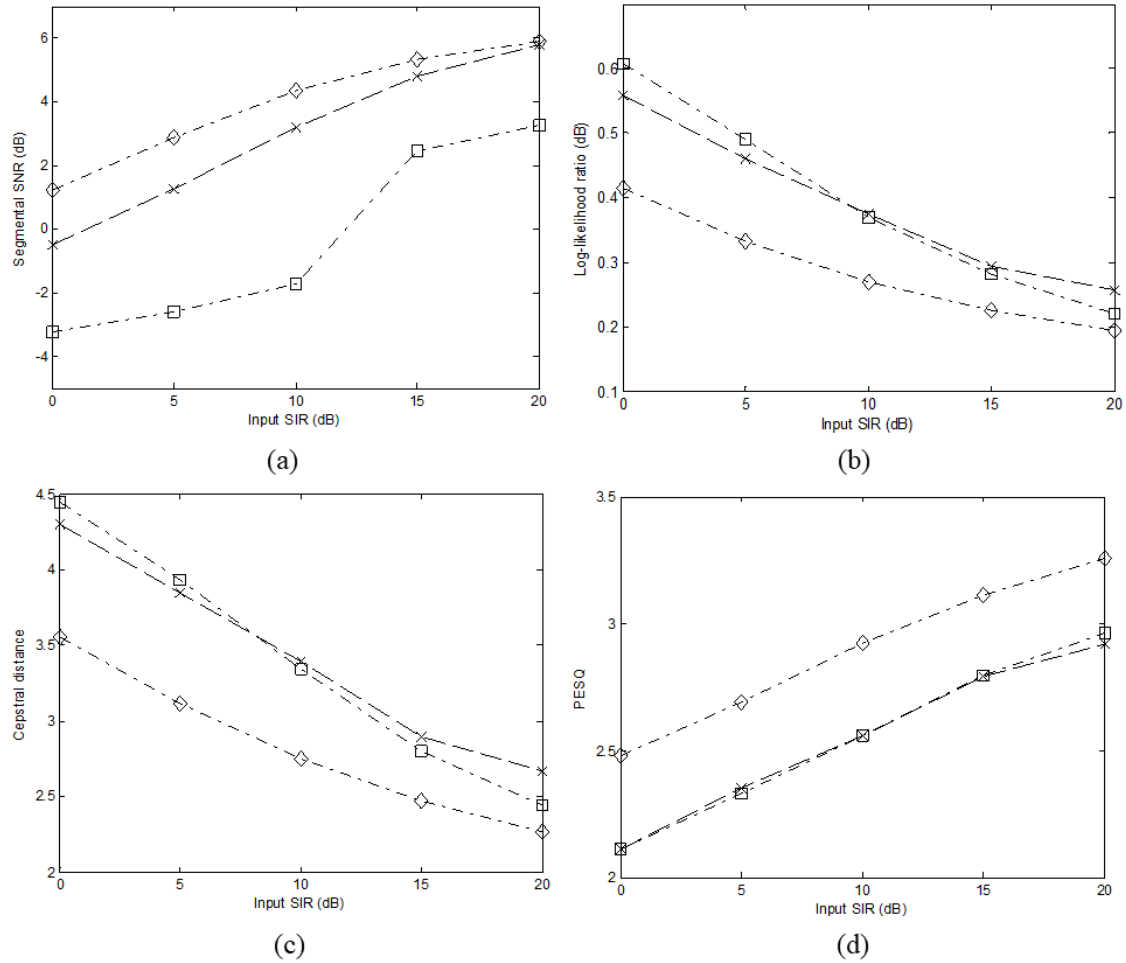


Fig. 3 Experimental results of 4 types of objective measures (a) segSNR (b) LLR (c) CD (d) PESQ. (-x-: noisy input, -□-: PMWF-M-MCRA, -◇-: PMWF-proposed)

와 방향성 간섭신호는 각각 배열 중심으로부터 20도와 -45도에 배치되었다.

변수내장형 다채널 위너필터 구현을 위해 μ 는 10으로 설정하였고 m_0 는 1번째 마이크로폰으로 설정하였다. 주파수 영역 필터를 시간영역으로 변환하고, [-64, 64]에 해당하는 필터 계수만 취하여 다시 주파수 영역으로 변환하는 일련의 과정을 거침으로써 비인과적 (noncausal) 유한 (finite) 임펄스 응답 제약 (constraint) 을 추가하였다. 기존 방법으로 다채널 최소 제한 재귀 평균 [12]을 구현하였다. 수식 (12)의 γ 와 ν 는 각각 10과 0.0775로 설정하였다.

4.2. 실험 결과 및 분석

제안한 방법을 평가하기 위해 segmental SNR (seg SNR), log-likelihood ratio (LLR), cepstral distance (CD), perceptual evaluation of speech quality (PESQ)를 측정[17]하였고 그림 3에 측정 결과를 비교하여 제시하였다. 그림 3(a)에서 기존 방법은 segSNR을 저하시킨 반면에 제안한 방법은 segSNR을 평균 1.024 dB 향상시켰다. 또한, 신호 왜곡도(LLR, CD)는 그림 3(b)와 그림 3(c)에서 유사한 경향으로 감소하는 것을 확인할 수 있다. 평균적으로 각각 0.101 dB와 0.590 만큼 신호 왜곡도가 향상 되었다. 특히, 그림 3(d)에서 평균적으로 0.345 정도의 현저한 PESQ 성능 향상을 보였다.

제안한 잡음추정기법의 성능향상은 목적신호의 방향성 정보를 사용함에서 기인한다. 다채널 최소 제어 재귀 평균 기법은 통계적으로 잘 정의된 잡음 추정 기법이지만 신호와 유사한 방향성 간섭신호가 존재하는 환경에서 목적신호의 방향성 정보 없이 에너지로만 잡음을 추정할 경우 정상동작 하지 않을 가능성이 크다. 실험 결과에서도 해당 환경에서 불충분한 잡음제거 결과를 보였다. 이 결과는 목적신호의 방향성 정보를 사용하지 않기 때문에 방향성 간섭신호를 목적신호로 오인(misperceive)하여 발생한다. 그러나 제안한 잡음추정 방법으로 변수내장형 다채널 위너필터를 산출하면 목적신호의 방향성 정보를 이용하여 각각의 부공간에서 목적신호 성분을 적절하게 산출하여 성공적으로 잡음 제거를 수행할 수 있다.

V. 결 론

본 논문에서는 목적신호의 방향정보를 이용한 새로운 잡음추정기법을 제안하였다. 제안한 방법은 잡음쉬인 마이크로폰 입력의 전력 스펙트럼 밀도 행렬의 고유값 분해하고 각각의 부공간에 해당하는 목적 신호 기여도 추정하는 방법이다. 제안한 잡음 추정 방법은 목적신호와 유사한 방향성 간섭신호가 존재할 경우에 강인하다. 하지만 제안한 지수 감소 함수에 기반한 가중치 할당 방법은 하나의 대안일 뿐, 최적이지 않기 때문에 최적의 가중치를 구하기 위한 연구는 향후에 수행할 계획이다.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2022R1G1A1008798).

REFERENCES

- [1] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multispeaker LCMV Beamformer and Postfilter for Source Separation and Noise Reduction," *IEEE Transaction Audio Speech Language Processing*, vol. 25, no. 5, pp. 940-951, May 2017.
- [2] Y. Kubo, T. Nakatani, M. Delcroix, K. Kinoshita, and S. Araki, "Mask-based MVDR Beamformer for Noisy Multisource Environments: Introduction of Time-varying Spatial Covariance Model," in *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, pp. 6855-6859, 2019.
- [3] P. Rakesh, S. S. Priyanka, and T. Kumar, "Performance evaluation of beamforming techniques for speech Enhancement," in *Proceedings of Fourth International Conference on Signal Processing Communication and Networking*, Chennai, India, pp. 1-5, 2017.
- [4] J. Park, J. Hong, J. Choi, and M. Hahn, "Determinant-based Generalized Sidelobe Canceller for Dual-Sensor Noise Reduction," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 8858-8868, May 2022.
- [5] S. M. Kim "Hearing Aid Speech Enhancement Using Phase Difference-Controlled Dual-Microphone Generalized Sidelobe Canceller," *IEEE Access*, vol. 7, no. 9, pp. 130663-130671, Sep. 2019.
- [6] J. Kim and M. Hahn, "Speech Enhancement Using a Two-Stage Network for an Efficient Boosting Strategy," *IEEE Signal Processing Letter*, vol. 26, no. 5, pp. 770 - 774, Mar. 2019.
- [7] J. Lee and H. G. Kang, "A Joint Learning Algorithm for Complex-Valued T-F Masks in Deep Learning-Based Single-Channel Speech Enhancement Systems," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 6, pp. 1098-1108, Jun. 2019.
- [8] D. Wang and J. Chen, "Supervised Speech Separation Based on Deep Learning: An Overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702-1726, Jun. 2018.
- [9] S. Pascual, A. Bonafonte, and J. Serra, "SEGAN: Speech Enhancement Generative Adversarial Network," in *Proceeding of Interspeech*, Stockholm, Sweden, pp. 3642-3646, 2017.
- [10] L. Zhang, M. Wang, Q. Zhang, X. Wang, and M. Liu, "PhaseDCN: A Phase-Enhanced Dual-Path Dilated Convolutional Network for Single Channel Speech Enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2561-2574, Jun. 2021.
- [11] S. Jeong and Y. Kim, "An Optimally-Modified Multichannel Wiener Filter Using Speech Presence Probability," *Smart Media Journal*, vol. 7, no. 3, pp. 9-15, Sep. 2018.
- [12] M. Souden, J. Chen, J. Benesty, and S. Affes, "An Integrated

- Solution for Online Multichannel Noise Tracking and Reduction,” *IEEE Transactions Audio, Speech, Language Processing*, vol. 19, no. 7, pp. 2159-2169, Sep. 2011.
- [13] I. Cohen, “Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging,” *IEEE Transactions Speech Audio Processing*, vol. 11, no. 5, pp. 466-475, Sep. 2003.
- [14] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, USA: Wiley, 1996.
- [15] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, “Speech enhancement based on the subspace method,” *IEEE Transaction Audio Speech Language Processing*, vol. 8, no. 5, pp. 497-507, Sep. 2000.
- [16] E. Warsitz and R. Haeb-Umbach, “Blind Acoustic Beamforming Based on Generalized Eigenvalue Eecomposition,” *IEEE Transaction Audio Speech Language Processing*, vol. 15, no. 5, pp. 1529-1539, Jul. 2007.
- [17] P. C. Loizou, *Speech Enhancement: Theory and Practice*, Boca Raton, FL, USA: CRC, 2007.
- [18] D. Pearce and H. Hirsch, “The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions,” in *Proceedings of Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000*, Beijing, China, pp. 16-20, 2000.
- [19] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, Apr. 1979.
- [20] E. A. P. Habets, “Generating sensor signals in isotropic noise fields,” *Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3464-3470, Dec. 2007.



홍정표(Jungpyo Hong)

2016년 한국과학기술원 전기및전자공학부 공학박사

2016년 ~ 2020년 국방과학연구소 선임연구원

2020년 ~ 현재 창원대학교 정보통신공학과 조교수

※관심분야: 마이크로폰 배열 기반 음성 향상, 소나 신호 처리, 압축 센싱, 능동 소음 제거, 인공 지능