

일반논문 (Regular Paper)

방송공학회논문지 제23권 제4호, 2018년 7월 (JBE Vol. 23, No. 4, July 2018)

<https://doi.org/10.5909/JBE.2018.23.4.519>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 마이크로폰 어레이 신호의 잡음 제거를 위한 강인한 다채널 위너 필터

정 준 영<sup>a)</sup>, 김 기 백<sup>a)†</sup>

### Robust Multi-channel Wiener Filter for Suppressing Noise in Microphone Array Signal

Junyoung Jung<sup>a)</sup>, Gibak Kim<sup>a)†</sup>

#### 요 약

본 논문에서는 다채널 위너 필터를 이용하여 마이크로폰 어레이에서 취득된 신호의 잡음을 제거하는 방법을 다룬다. 다채널 위너 필터는 음성 신호의 방향에 대한 정보를 필요로 하지 않는 필터로서 단일 음성 음원의 경우, 음성 왜곡을 발생시키지 않는 MVDR (Minimum Variance Distortionless Response) 공간 필터와 단일 채널 스펙트럼 필터로 분리될 수 있다. MVDR의 방향벡터에 해당하는 단일 음성 음원과 마이크로폰 어레이 간의 음향 전달 함수는 다채널 위너 필터의 부공간 분해 (subspace decomposition)를 이용하여 추정할 수 있다. 이 때 상관 행렬 추정 과정에서 발생하는 오차로 인해 추정되는 음향 전달 함수에도 오차가 발생하게 되며 이에 따라 다채널 위너 필터를 구성하는 MVDR은 음성 왜곡을 발생시키게 된다. 이러한 음성 왜곡을 완화시키기 위해 diagonal loading을 적용하고 실험을 통해 그 효과를 검증한다. 실험에서는 7개의 선형 마이크로폰으로 수집된 데이터를 이용하였으며 잡음을 섞기 전 신호와 잡음을 섞은 후 필터를 통과시킨 신호 간의 MFCC 오차를 측정한다. 실험 결과, diagonal loading을 통해 MFCC 오차를 줄일 수 있음을 확인하였다.

#### Abstract

This paper deals with noise suppression of multi-channel data captured by microphone array using multi-channel Wiener filter. Multi-channel Wiener filter does not rely on information about the direction of the target speech and can be partitioned into an MVDR (Minimum Variance Distortionless Response) spatial filter and a single channel spectral filter. The acoustic transfer function between the single speech source and microphones can be estimated by subspace decomposition of multi-channel Wiener filter. The errors are incurred in the estimation of the acoustic transfer function due to the errors in the estimation of correlation matrices, which in turn results in speech distortion in the MVDR filter. To alleviate the speech distortion in the MVDR filter, diagonal loading is applied. In the experiments, database with seven microphones was used and MFCC distance was measured to demonstrate the effectiveness of the diagonal loading.

Keyword : multi-channel Wiener filter, microphone array, noise suppression, beamforming

a) 숭실대학교 전기공학부(School of Electrical Engineering, Soongsil University)

† Corresponding Author : 김기백(Gibak Kim)

E-mail: imkgb27@ssu.ac.kr

Tel: +82-2-828-7266

ORCID:<https://orcid.org/0000-0001-5114-4117>

※ 이 논문은 한국 산업통상자원부의 로봇산업융합핵심기술사업 프로그램 No.10048474, ‘고령화 세대에게 개인별 특화된 복지 서비스를 제공하기 위한 빅데이터 기반의 서비스 로봇개발’의 지원으로 수행되었음.

※ This material is based upon work supported by the Ministry of Trade, Industry&Energy(MOTIE, Korea) under Industrial Technology Innovation Program No.10048474, ‘Development of a Service Robot based on Big Data for Providing Aging Generation with personalized Welfare Services’

· Manuscript received May 8, 2018; Revised July 10, 2018; Accepted July 10, 2018.

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

## 1. 서론

잡음 환경에서 취득된 음성 신호의 잡음 제거는 실제 환경에서 음성 통신이나 음성 인식을 위하여 요구되는 주요 기술 중 하나이다. 실제 환경에서는 다양한 잡음에 노출되고, 경우에 따라서는 심각한 성능 저하를 초래하기도 한다. 하드웨어의 비약적인 발전에 힘입어 두 개 이상의 마이크가 설치된 기기가 늘어나고 있으며, 이에 따라 빔포밍 등 다채널 필터를 이용하여 음성 왜곡을 최소화하면서 높은 잡음 제거 성능을 보이는 알고리즘들이 많이 연구되고 있다. 단일 채널 신호가 시간-주파수 영역의 정보만을 이용하는 반면, 다채널 신호는 신호의 공간적인 정보도 갖고 있기 때문에 이를 이용하면 음성 왜곡을 최소화할 수 있다. 마이크로폰 어레이로부터 취득된 신호의 잡음 제거 알고리즘으로는 고정 빔포밍, 적응 빔포밍, 포스트 필터링, 다채널 위너 필터링(Multi-channel Wiener Filter: MWF) 등이 적용되고 있다<sup>[1]</sup>.

MWF는 다른 빔포밍 알고리즘들과는 달리 신호의 방향 정보를 사용하지 않고 음성 성분과 잡음 성분의 상관 행렬(correlation matrix)을 추정하고 이로부터 필터를 추정하므로 방향 감지를 필요로 하지 않는다. 잡음이 심한 환경에서는 신호의 방향을 추정하기가 어렵고, 반향(reverberation)이 심한 환경에서는 방향 벡터만으로는 신호의 채널 특성을 나타내기가 어렵다. MWF는 방향 감지나 그에 따른 방향 벡터를 사용하지 않으므로 위와 같은 문제로부터 자유

롭다. 보다 효과적인 MWF의 구현을 위해 부공간 분해(subspace decomposition)를 이용하는 연구들이 진행되어 왔다<sup>[2-9]</sup>. 부공간 분해를 이용하는 구현에서는 다채널 신호를 부공간으로 분해한 후, 부공간으로 신호들을 투영(projection)하여 공간 필터링을 하고 스펙트럴 필터링으로 추가 잡음을 제거한다(그림 1). 만일 주파수 영역에서 고유값 분해(eigenvalue decomposition)를 통해 부공간 분해가 이루어진다면 가장 큰 고유값에 해당하는 고유 벡터(eigenvector)가 단일 음성 음원과 마이크로폰 어레이 간의 음향 전달 함수에 해당하게 된다.

단일 음성 음원의 경우, MWF는 MVDR (Minimum Variance Distortionless Response) 빔포머와 단일 채널 스펙트럴 필터로 분해될 수 있다. 즉, 음성 신호의 방향 벡터(steering vector)를 이용하는 MVDR 빔포머를 적용하여 음성의 왜곡없이 잡음을 제거하고, 단일 채널 위너 필터를 이용하여 잔류 잡음을 제거하게 된다. MVDR 빔포머의 방향 벡터는 단일 음성 음원과 마이크로폰 어레이 간의 음향 전달 함수로 표현되고, 이것은 부공간 분해를 통해 구한 고유 벡터로 대체될 수 있다. 그러나 부공간 분해를 통한 고유 벡터 추정시, 상관 행렬 추정 오차 등으로 인해 오차가 발생하게 되고 이로 말미암아 MVDR 빔포머를 통과한 신호는 음성 왜곡을 포함하게 된다.

본 논문에서는 MWF를 MVDR과 스펙트럴 필터로 분해하여 구현한다. 또한 고유값 분해를 통해 MWF를 부공간에서 구현하여 음향 전달 함수를 추정하고, 추정된 음향 전달 함수를 이용하여 분리된 MVDR을 구현한다. MVDR이 음향 전달 함수 오차에 강인한 특성을 갖도록 diagonal loading을 적용한다. Diagonal loading 적용에 따른 MWF의 성능을 평가하기 위해 7개의 마이크로 구성된 마이크로폰 어레이 데이터를 이용하여 MFCC(Mel-Frequency Cepstral Coefficient) 오차를 측정한다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 MWF와 MWF의 MVDR, 스펙트럴 필터 분해에 대해 설명하고, 3장에서는 부공간 분해를 이용한 음향 전달 함수의 추정과 오차에 강인한 MVDR의 구현을 위한 diagonal loading에 대해 설명한다. 마이크로폰 어레이 데이터에 대한 MFCC 오차 테스트를 포함하는 실험결과는 4장에서 제시한다.

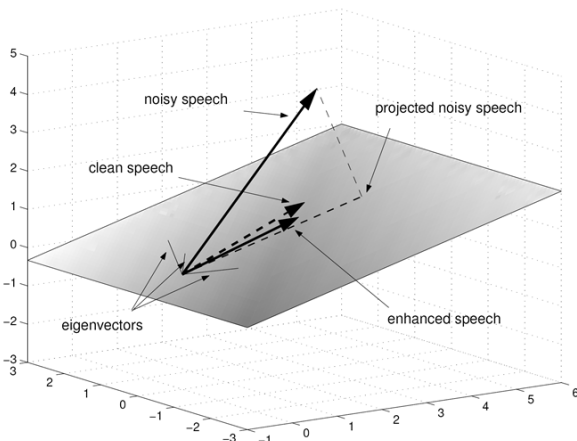


그림 1. 부공간 분해를 이용한 잡음제거  
Fig. 1. Noise reduction using subspace decomposition

## II. 다채널 위너 필터 (MWF)

본 장에서는 마이크로폰 어레이를 이용하여 잡음을 제거하는 방법으로서, 최소 평균 자승 오차 (Minimum Mean Squared Error: MMSE) 추정에 근거한 다채널 위너 필터 (MWF)와 MWF의 MVDR, 스펙트럴 필터 분리에 대하여 설명한다.

### 1. MMSE에 근거한 MWF

먼저  $M$ 개의 마이크로폰으로부터 다채널 신호는 다음 식과 같이 주파수 영역에서 음성 신호와 잡음 성분으로 분해할 수 있다.

$$\mathbf{y}(f) = \mathbf{x}(f) + \mathbf{n}(f) \quad (1)$$

음성이 하나의 음원이라고 가정하면 다채널 음성 신호  $\mathbf{x}(f)$ 는 단일 음성 음원  $S(f)$ 와 다채널 음향 전달 함수 (acoustic transfer function)  $\mathbf{h}(f)$ 를 고려하여 다음과 같이 표시할 수 있다.

$$\mathbf{x}(f) = S(f)\mathbf{h}(f) \quad (2)$$

여기서  $\mathbf{y}$ ,  $\mathbf{x}$ ,  $\mathbf{h}$ ,  $\mathbf{n}$ 은 모두  $M$ 차원 벡터이다.

$M$ 차원 잡음 제거 필터를  $\mathbf{w}$ 라고 하면 필터를 통과한 신호  $Z$ 는 다음과 같다.

$$Z(f) = \mathbf{w}^H(f)\mathbf{y}(f) \quad (3)$$

수식을 간단하게 표현하기 위해 주파수를 나타내는 ( $f$ )는 이후부터 생략하기로 한다. 필터를 통과한 출력 신호  $Z$ 와 첫 번째 마이크의 음성 성분 간의 평균 자승 오차를 최소화하는 MWF는 다음과 같이 주어진다.

$$\mathbf{w} = \mathbf{R}_y^{-1}\mathbf{R}_x\mathbf{e}_1 \approx \mathbf{R}_y^{-1}(\mathbf{R}_y - \mathbf{R}_n)\mathbf{e}_1 \quad (4)$$

여기서  $\mathbf{R}_y$ ,  $\mathbf{R}_x$ ,  $\mathbf{R}_n$ 은  $M \times M$  정방 행렬 (square matrix)로서 각각  $\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H]$ ,  $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H]$ ,  $\mathbf{R}_n = E[\mathbf{n}\mathbf{n}^H]$ 으로

나타나는 상관 행렬 (correlation matrix)이다. 그리고  $\mathbf{e}_1$ 은  $M$ 차원 벡터로서 첫 번째 원소만 1이고 나머지 원소들은 0이다.

### 2. MWF의 분해

단일 음성 음원의 경우 식(2)와 같이 나타나고 다채널 신호의 음성 성분  $\mathbf{x}$ 의 상관 행렬은 다음과 같이 표현된다.

$$\mathbf{R}_x = \mathbf{h}P_S\mathbf{h}^H, P_S = E[S^2] \quad (5)$$

$\mathbf{R}_y$ 를 다음 식과 같이 근사화하고,

$$\mathbf{R}_y \approx \mathbf{R}_x + \mathbf{R}_n \quad (6)$$

식(5)와 식(6)을 식(4)에 대입하면 다음 결과를 얻게 된다. 이 때, 다채널 음향 전달 함수  $\mathbf{h}$ 의 첫 번째 원소는 1로 가정하였다.

$$\mathbf{w} = (\mathbf{h}P_S\mathbf{h}^H + \mathbf{R}_n)^{-1}\mathbf{h}P_S \quad (7)$$

위의 식을 Woodbury의 정리 (matrix inversion lemma)<sup>[9]</sup>를 이용하여 정리하면 다음 식과 같이 쓸 수 있다.

$$\mathbf{w} = \mathbf{R}_n^{-1}\mathbf{h}P_S - \mathbf{R}_n^{-1}P_S\mathbf{h}(1 + \mathbf{h}^H\mathbf{R}_n^{-1}P_S\mathbf{h})^{-1}\mathbf{h}^H\mathbf{R}_n^{-1}\mathbf{h}P_S \quad (8)$$

아래의 식을 가정하면

$$\sigma_n = (\mathbf{h}^H\mathbf{R}_n^{-1}\mathbf{h})^{-1} \quad (9)$$

식(8)은 다음과 같이 정리할 수 있다.

$$\begin{aligned} \mathbf{w} &= \frac{P_S}{P_S + \sigma_n} \sigma_n \mathbf{R}_n^{-1} \mathbf{h} = w_p \mathbf{w}_{MVDR} \\ w_p &= \frac{P_S}{P_S + \sigma_n}, \mathbf{w}_{MVDR} = \sigma_n \mathbf{R}_n^{-1} \mathbf{h} \end{aligned} \quad (10)$$

식(10)은 두 개의 필터가 순차적으로 적용된 것으로 이해할 수 있는데, 첫 번째 필터  $\mathbf{w}_{MVDR}$ 은 다채널 음향 전달 함수  $\mathbf{h}$ 를 반향 (reverberation) 성분을 고려하지 않은 방향

벡터로 가정한다면 MVDR 이 되고 두 번째 필터  $w_p$  는 MVDR 필터의 출력에 적용되는 단일 채널 위너 필터이다 (그림 2).

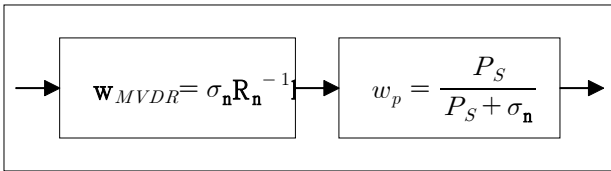


그림 2. MVDR필터와 스펙트럴 필터로 분리된 MWF  
Fig. 2. MWF partitioned into MVDR and spectral filters

### III. MWF의 부공간 분해와 MVDR의 diagonal loading

#### 1. MWF의 부공간 분해

일반화된 고유값 분해 (Generalized EigenValue Decomposition: GEVD)<sup>[11]</sup>를 적용하여 다음과 같이 상관 행렬  $R_y$  과  $R_n$  을 동시에 대각행렬( $A_y, A_n$ )로 변환한다.

$$Q^H R_y Q = A_y, \quad Q^H R_n Q = A_n \quad (11)$$

위의 식으로부터 상관 행렬들을 아래와 같이 표현할 수 있다.

$$R_y = \bar{Q} A_y \bar{Q}^H, \quad R_n = \bar{Q} A_n \bar{Q}^H, \quad \bar{Q} = Q^{-H} \quad (12)$$

식(12)를 식(4)에 대입하여 아래의 식을 얻는다.

$$w = Q(I - A_y^{-1} A_n) \bar{Q}^H e_1 \quad (13)$$

잡음과 음성 신호의 상관 관계를 무시하면  $R_x$  는  $R_y$  과  $R_n$  의 차로 표현될 수 있고, 식(12)를 적용하면  $R_x$  는 아래 식으로 표현할 수 있다.

$$R_x \approx \bar{Q}(A_y - A_n) \bar{Q}^H \quad (14)$$

다채널 신호의 음성 성분  $x$  는 식(2)와 같이 단일 채널 음성  $S$ 와 음향 전달 함수의 곱으로 표현되므로 행렬  $R_x$  의 rank는 1이고, 이를 적용하면 식(14)는 아래의 식과 같이 나타낼 수 있다.

$$R_x \approx (\lambda_{y,1} - \lambda_{n,1}) \bar{q}_1 \bar{q}_1^H \quad (15)$$

여기서  $\bar{q}_1$  은  $\bar{Q}$  의 첫 번째 열벡터이고,  $\lambda_{y,1}$  과  $\lambda_{n,1}$  은  $A_y$  와  $A_n$  의 첫 번째 원소이다. 식(15)를 식(5)와 비교하면 다음과 같이 음향 전달 함수  $h$  와 단일 음성 음원의 power spectral density  $P_S$  를 다음과 같이 추정할 수 있다.

$$h = \bar{q}_1, \quad P_S = \lambda_{y,1} - \lambda_{n,1} \quad (16)$$

식(12)로부터  $R_n^{-1} = \bar{Q}^{-H} A_n^{-1} \bar{Q}^{-1}$  을 유도하고,  $\bar{q}_1 = \bar{Q} e_1$  임을 이용하면 식(10)에 있는  $\sigma_n$  이  $\lambda_{n,1}$  과 같음을 알 수 있다. 따라서 식(10)의 필터  $w$  는 다음과 같이 나타낼 수 있다.

$$w = \frac{\lambda_{y,1} - \lambda_{n,1}}{\lambda_{y,1}} \lambda_{n,1} R_n^{-1} \bar{q}_1 \quad (17)$$

#### 2. MVDR의 diagonal loading

단일 음성 음원의 경우는 MWF가 식(10)과 같이 다채널 공간 필터 MVDR과 단일 채널 스펙트럴 필터로 나뉘어질 수 있다. 여기서 MVDR은 이론상 음성 음원을 왜곡시키지 않는 필터이나 방향 벡터에 오차가 있다면 왜곡이 발생하게 된다. MWF에서  $\bar{q}_1$  으로  $h$  를 추정하는 과정에서 발생할 수 있는 오차는 다음과 같다.

##### 1.1 상관행렬( $R_y, R_n$ ) 추정 오차

마이크로폰 어레이로부터 취득되는 데이터의 상관 행렬  $R_y$  는 관찰되는 데이터 벡터의 외적 (outer product:  $yy^H$ ) 들을 누적시키거나 망각 인자 (forgetting factor)를 이용하여 구할 수 있다. 잡음 성분의 상관 행렬  $R_n$  은 잡음만 존재

하는 구간에서 추정할 수 있으며,  $\mathbf{R}_y$  와 비슷한 방법으로 추정할 수 있다.

### 1.2 $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n$ 근사화에 따른 오차

이러한 가정은 잡음과 음성 성분의 교차 상관값 (cross correlation)을 무시할 수 있다는 가정이 뒷받침되어 있다. 즉, 음성 성분의 자기 상관값의 크기가 음성과 잡음의 교차 상관값의 크기보다 훨씬 크다는 가정 ( $|E[x_i x_j^*]| \gg |E[x_i n_j^*]|$ )을 전제하고 있다. 그러나 신호 대 잡음 비(Signal-to-Noise Ratio: SNR)가 낮은 경우는 식(6)과 식(14)에서 발생하는 오차가 무시할 수 없는 수준이 된다.

위와 같은 오차로 말미암아 식(10)의 MVDR 필터에 의한 결과는 음성 왜곡을 필연적으로 수반하게 되는데 이를 최소화하기 위한 방법 중 하나로서 diagonal loading이 널리 사용되고 있다<sup>[12]</sup>. MVDR 필터 계수의 norm이 커질수록 방향 벡터와 관련된 오차가 MVDR 공간 응답 (spatial response)에 미치는 영향이 커진다는 사실에 기반하여, MVDR 필터를 추정할 때 필터 계수의 2차 norm ( $\|\mathbf{w}\|^2$ )을 특정값 이하로 제한하는 조건을 추가하여 오차에 따른 음성 왜곡을 줄일 수 있다. 이러한 제한 조건을 추가하여 MVDR 필터를 유도하면 잡음 상관 행렬  $\mathbf{R}_n$  대신  $\mathbf{R}_n + \mu\mathbf{I}$ , 즉 대각 행렬이 추가된 형태로 바뀌게 된다. 이것을 diagonal loading이라 부른다. 본 논문에서는 diagonal loading을 적용하여 방향 벡터 추정 오차에 따른 음성 왜곡을 줄이도록 유도된 필터를 사용한다. 필터 계수의 Diagonal loading은 고유값 분해에서 고유값이 너무 작아서 발생하는 numerical 문제에 대한 해결책으로 사용되기도 한다. Diagonal loading을 적용하면 식(17)필터는 다음 식과 같이 된다.

$$\mathbf{w} = \frac{\lambda_{y,1} - \lambda_{n,1}}{\lambda_{y,1}} \lambda_{n,1} (\mathbf{R}_n + \mu\mathbf{I})^{-1} \mathbf{q}_1 \quad (18)$$

## IV. 실험 결과

### 1. 실험 환경 및 실험 과정

필터의 성능 검증을 위해 백색 잡음과 음성 간섭 잡음을

사용하였다. 음성 간섭 잡음으로는 16초간 뉴스 기사를 읽는 음성을 사용하였다. 목적 음성은 TIDIGITS 데이터베이스에 포함된 연결 숫자음 발성을 사용하였다<sup>[13]</sup>. TIDIGITS 데이터베이스는 20kHz의 샘플링 주파수를 갖는 데이터인데 16kHz로 다운샘플링하였다.

실험을 위한 다채널 데이터 생성을 위해 RWCP Sound Scene Database<sup>[14]</sup>에 포함된 임펄스 응답을 사용하였다. 다채널 임펄스 응답과 음원을 컨벌루션하여 멀티 채널 데이터를 생성하였다. 임펄스 응답의 반향 시간 (reverberation time)은 300ms이다. 그림 3과 같이 목적 신호와 간섭 신호는 마이크로폰 어레이의 중심으로부터 2m 떨어져 있고 40°의 각도를 이루고 있다. 7개의 마이크가 서로 선형으로 배치되어 있으며, 간격은 5.66cm이다. 신호 대 잡음 비는 -5, 0, 5, 10dB가 되도록 하였다.

입력 데이터는 32ms(=512샘플)가 하나의 프레임을 이루도록 하였고, 프레임 간 오버랩은 50%가 되도록 하면서 프레임을 이동하였다. 각 프레임 데이터는 Hanning 윈도우가 적용된 후 FFT(Fast Fourier Transform)을 수행하였다.

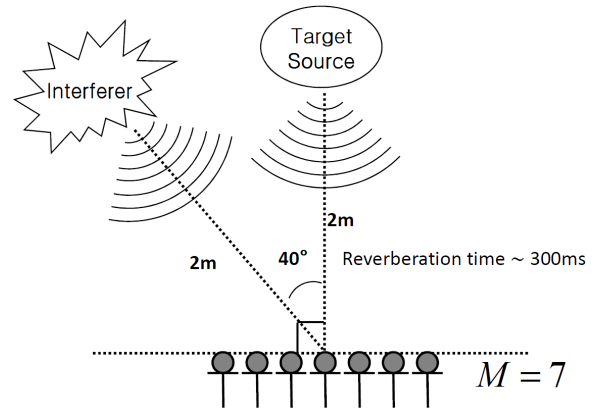


그림 3. 마이크로폰 어레이 실험 환경  
 Fig. 3. Testing environment with a linear microphone array

### 2. 실험 결과

간섭 신호를 섞지 않은 마이크로폰 어레이의 첫 번째 마이크 신호와 간섭 신호가 섞여진 어레이 신호를 필터에 통과시켜 얻은 신호 간의 MFCC 거리를 측정하여 필터의 성능을 평가하였다. MFCC는 음성 신호처리에서 가장 널리

사용되는 대표적인 특징으로서 두 신호 간의 거리 평가를 통해 두 신호의 차이를 음성 신호 측면에서 평가할 수 있다. MFCC 거리는 다음 식으로 구하였다.

$$\text{MFCC distance} = \frac{1}{K} \sum_{k=1}^K \sqrt{\sum_i [c_{x_1}^i(k) - c_z^i(k)]^2} \quad (19)$$

여기서  $c_{x_1}^i(k)$ 와  $c_z^i(k)$ 는 각각 첫 번째 마이크의 음성 성분과 필터 출력 신호의  $k$ 번째 프레임 데이터의  $i$ 번째 MFCC이고,  $K$ 는 데이터의 프레임 개수이다.

그림 4에 3가지 필터에 대한 MFCC 거리 측정 결과를 나타내었다. Filter\_1은 식(17)에 유도된 필터로서 MWF를

부공간 분해하고,  $\mathbf{R}_x$ 의 rank가 1임을 이용하여 부공간 벡터 가장 큰 고유값을 갖는 벡터 공간만 이용하여 구현한 필터이다. Filter\_2는 MWF에서 분리된 MVDR에 diagonal loading을 적용한 필터(식(18))이다. 비교를 위해서 잡음 제거 필터를 통과하지 않은 신호의 MFCC 거리도 함께 나타내었다(Unprocessed).

Diagonal loading에서  $\mu$  값은 다양한 방법으로 계산할 수 있으나<sup>[15]</sup>, 일반적으로 잡음 상관 행렬의 대각 원소들의 합을 마이크로폰의 수( $M$ )로 나눈 값에 비례하여 정할 수 있다<sup>[16]</sup>. 본 실험에서는 아래 식과 같이  $\mu$  값을 정하였다. 0.01은 실험을 통해 선택된 값이다.

$$\mu = 0.01 \text{Trace}(\mathbf{R}_n) / M \quad (20)$$

그림 4의 결과에서 보듯이 diagonal loading을 적용하여 추가적인 성능 향상을 도모할 수 있음을 알 수 있다. 특히 낮은 SNR에서의 향상이 두드러지는데, 그 이유는 낮은 SNR에서 방향 벡터(음향 전달 함수)의 오차가 상대적으로 크게 발생하기 때문으로 추정할 수 있다.

## V. 결론

본 논문에서는 마이크로폰 어레이로 수집된 잡음 환경 음성의 잡음을 제거하기 위해 다채널 위너 필터(MWF)를 적용하였다. MWF는 공간 필터인 MVDR과 스펙트럴 필터인 단일 채널 위너 필터로 분리될 수 있는데, 각 필터들은 MWF의 부공간 분해를 통해 구현될 수 있다. 부공간 분해 과정에서 추정되는 음향 전달 함수는 MVDR에서 사용될 수 있는데, 음향 전달 함수의 추정 오차로 인해 MVDR에 음성 왜곡이 발생하게 된다. 이를 완화하기 위해 대표적인 diagonal loading 방법을 사용하였다. 마이크로폰 어레이 임펄스 응답을 이용하여 다채널 음성 신호를 생성하고 백색 잡음 및 음성 간섭 잡음 환경에 대해 MFCC오차를 실험하였다. 실험 결과 diagonal loading을 적용하여 MWF의 잡음 제거 성능을 향상시킬 수 있음을 확인하였다.

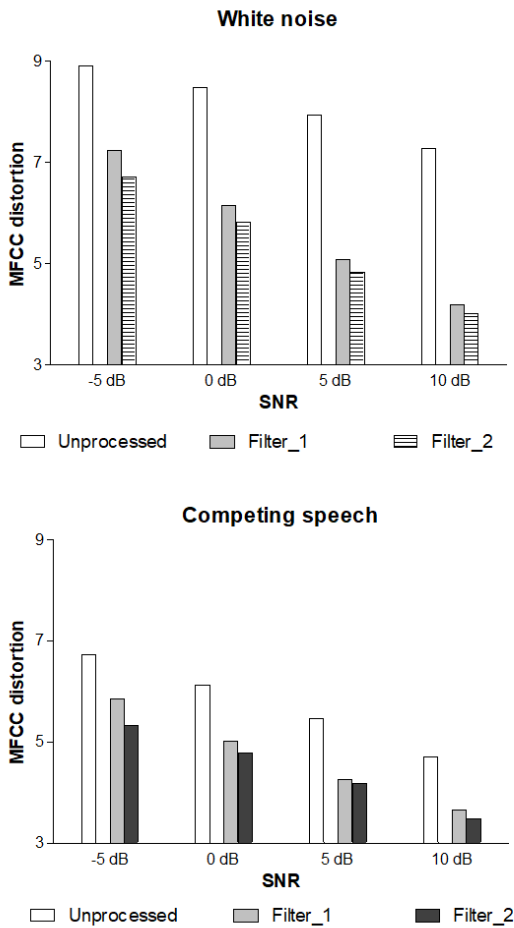


그림 4. 필터의 잡음 제거 성능: MFCC 거리 측정  
Fig. 4. Results of filters: MFCC distortion

## 참 고 문 헌 (References)

- [1] M. Brandstein and D. Ward (Eds.), *Microphone Arrays*, Springer-Verlag, 2001. <https://doi.org/10.1007/978-3-662-04619-7>
- [2] S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230-2244, Sep. 2002. [https://doi.org/10.1007/978-3-662-04619-7\\_6](https://doi.org/10.1007/978-3-662-04619-7_6)
- [3] A. Spriet, M. Moonen, and J. Wouters, "Robustness Analysis of Multichannel Wiener Filtering and Generalized Sidelobe Cancellation for Multimicrophone Noise Reduction in Hearing Aid Applications," *IEEE Transactions Speech Audio Processing*, vol. 13, no. 4, July 2005, pp. 487-503. <https://doi.org/10.1109/tsa.2005.845821>
- [4] S. Doclo and M. Moonen, "Combined Frequency-Domain Dereverberation and Noise Reduction Technique for Multi-microphone Speech Enhancement," *Int. Workshop Acoustic Echo Noise Control*, Darmstadt, Germany, Sept. 2001, pp. 31-34.
- [5] W. Herboldt, *Sound Capture for Human/Machine Interfaces*, Springer-Verlag, 2005.
- [6] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 260-276, 2010 <https://doi.org/10.1109/tasl.2009.2025790>
- [7] Gibak Kim, "SNR-based weight control for the spatially preprocessed speech distortion weighted multi-channel Wiener filtering," *Journal of Broadcast Engineering*, Vol. 18, No.3, pp.455-462, March, 2013. <https://doi.org/10.5909/jbe.2013.18.3.455>
- [8] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 785-799, 2014. <https://doi.org/10.1109/taslp.2014.2304240>
- [9] Z. Wang, "Rank-1 constrained Multichannel Wiener Filter for speech recognition in noisy environments," *Computer Speech & Language*, vol. 49, pp. 37-51, 2018. <https://doi.org/10.1016/j.csl.2017.11.003>
- [10] D. Tylavsky and G. Sohie, "Generalization of the matrix inversion lemma," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1050-1052, 1986. <https://doi.org/10.1109/proc.1986.13587>
- [11] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition, 1996. <https://doi.org/10.2307/3619868>
- [12] H. Van Trees, *Detection, Estimation and Modulation Theory, Part IV: Optimum Array Processing*, New York: Wiley, 2002. <https://doi.org/10.1002/0471221104>
- [13] R. Leonard, "A database for speaker-independent digit recognition," *Proceedings of ICASSP*, pp. 111-114, 1984. <https://doi.org/10.1109/icassp.1984.1172716>
- [14] "RWCP Sound Scene Database in Real Acoustical Environments," *Real World Computing Partnership*, ©1998-2001
- [15] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 449-458, 2010. <https://doi.org/10.1109/icassp.2008.4518112>
- [16] Y. Xiao, J. Yin, H. Qi, H. Yin, and G. Hua, "MVDR algorithm based on estimated diagonal loading for beamforming," *Hindawi Mathematical Problems in Engineering*, pp. 1-7, 2017. <https://doi.org/10.1155/2017/7904356>

---

## 저 자 소 개

---

### 정 준 영



- 2017년 : 숭실대학교 전기공학부 학사
- 2017년 ~ 현재 : 숭실대학교 전기공학부 석사과정
- ORCID : <https://orcid.org/0000-0002-8327-7181>
- 관심분야 : 머신러닝 및 딥러닝, 음성신호처리, 전력신호처리

### 김 기 백



- 1994년 : 서울대학교 전자공학과 학사
- 1996년 : 서울대학교 전자공학과 석사
- 2007년 : 서울대학교 전기컴퓨터공학부 박사
- 1996년 ~ 2000년 : LG전자기술원 연구원
- 2000년 ~ 2003년 : (주)보이스웨어 선임연구원
- 2008년 ~ 2010년 : Univ. of Texas at Dallas, Research Associate
- 2011년 ~ 현재 : 숭실대학교 전기공학부 교수
- ORCID : <https://orcid.org/0000-0001-5114-4117>
- 관심분야 : 머신러닝 및 딥러닝, 음성신호처리, 전력신호처리