한국음향학회지 제39권 제5호 pp. 490~496 (2020)
The Journal of the Acoustical Society of Korea Vol.39, No.5 (2020)
https://doi.org/10.7776/ASK.2020.39.5.490

pISSN : 1225-4428
eISSN : 2287-3775

다채널 위너 필터의 주성분 부공간 벡터 보정을 통한 잡음 제거 성능 개선

Improved speech enhancement of multi-channel Wiener filter using adjustment of principal subspace vector

김기백[1†]

(Gibak Kim[1†])

[1]숭실대학교 전기공학부

(Received July 31, 2020; revised September 4, 2020; accepted September 9, 2020)

초 록: 본 논문에서는 잡음 환경에서 다채널 위너 필터의 성능을 향상시키기 위한 방법을 제안한다. 부공간(subspace) 기반의 다채널 위너 필터를 설계하는 경우, 목적 신호가 단일 음원인 경우는 음성 상관 행렬의 주성분 부공간에서 음성 성분을 추정할 수 있다. 이 때, 음성 상관 행렬은 음성과 간섭 잡음의 교차 상관도가 음성 상관 행렬에 비해 무시할만한 수준이라는 가정하에 신호 상관 행렬에서 간섭 잡음의 상관 행렬을 차감하여 추정하게 된다. 그러나 간섭 잡음 수준이 높아지게 되면 이러한 가정이 더 이상 유효하지 않게 되며 이에 따라 주성분 부공간 추정 오차도 증가하게 된다. 본 연구에서는 음성 존재 확률과 목적 신호의 방향 벡터를 이용하여 주성분 부공간을 보정하는 방법을 제안한다. 주성분 부공간에서 다채널 음성 존재 확률을 유도하고 주성분 부공간 벡터를 보정하는데 적용하였다. 실험을 통해 제안하는 방법이 잡음 환경에서 다채널 위너 필터의 성능을 향상시키는 것을 확인할 수 있다.
핵심용어: 잡음 제거, 다채널 위너 필터, 음성 존재 확률, 부공간 분해

ABSTRACT: We present a method to improve the performance of the multi-channel Wiener filter in noisy environment. To build subspace-based multi-channel Wiener filter, in the case of single target source, the target speech component can be effectively estimated in the principal subspace of speech correlation matrix. The speech correlation matrix can be estimated by subtracting noise correlation matrix from signal correlation matrix based on the assumption that the cross-correlation between speech and interfering noise is negligible compared with speech correlation. However, this assumption is not valid in the presence of strong interfering noise and significant error can be induced in the principal subspace accordingly. In this paper, we propose to adjust the principal subspace vector using speech presence probability and the steering vector for the desired speech source. The multi-channel speech presence probability is derived in the principal subspace and applied to adjust the principal subspace vector. Simulation results show that the proposed method improves the performance of multi-channel Wiener filter in noisy environment.
Keywords: Noise reduction, Multi-channel Wiener filter, Speech presence probability, Subspace decomposition

PACS numbers: 43.72.Dv, 43.60.Fg

†**Corresponding author:** Gibak Kim (imkgb27@ssu.ac.kr)
School of Electrical Engineering, Soongsil University, 369 Sangdoro, Dongjak-gu, Seoul 06978, Republic of Korea
(Tel: 82-2-828-7266)

Copyright ⓒ2020 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서 론

잡음 제거 기술은 음성 신호 처리 분야의 주요 기술 중 하나로서 잡음이 존재하는 실생활 환경에서 자동 음성 인식, 음성 통신의 성능 보장을 위해 요구되는 기술이다. 하드웨어의 발전에 따라 두 개 이상의 마이크를 이용하는 기기들이 증가하고 있으며, 이에 따라 다채널 잡음 제거 기술에 관한 연구도 확대되어 왔다.

단일 채널 신호에 비해 다채널 신호에는 신호의 공간 정보가 포함되어 있어 음성 왜곡을 최소화하면서 잡음 제거를 효과적으로 수행할 수 있다. 다채널 잡음 제거에 관한 초기 연구에서는 전통적인 빔포밍 기법을 활용하는 연구가 주로 수행되었고, 이후에는 음성 신호와 출력 신호 간의 자승 오차를 통계적 평균 관점에서 최소화하는 필터인 다채널 위너 필터(Multi-channel Wiener Filter, MWF)에 관한 연구가 활발히 진행되었다.[1-7]

다채널 위너 필터를 구하기 위해서는 음성과 잡음의 상관 행렬들을 추정해야 하는데, 이들 상관 행렬 추정의 정확도에 따라 필터의 잡음 제거 성능이 좌우된다. 음성과 잡음의 상관 행렬들을 추정할 때, 음성 존재 확률을 이용하여 잡음과 음성의 상관 행렬을 추정할 수 있다.[7-9] 다채널 신호를 이용한 음성 존재 확률은 단일 채널 신호를 사용했을 때보다 우수한 성능을 보이며, 상관 행렬 추정뿐만 아니라 다채널 위너 필터의 음성 왜곡 가중치를 설정하는데도 효과적으로 사용되었다.[2,3,10]

본 논문에서는 다채널 신호의 음성 존재 확률을 이용하여 다채널 위너 필터를 보정하는 방법을 제안하고자 한다. 부공간(subspace)에서 필터를 구할 때, 하나의 음원으로부터 음성이 여러 마이크에 도달하는 경우는 음성 상관 행렬의 랭크(rank)는 1이 되고 음성 상관 행렬의 주성분 벡터만을 이용하여 다채널 위너 필터의 성능을 향상시킬 수 있다.[1,5,6] 이 때 주성분 벡터는 이론적으로 음성 신호의 음향 전달 함수와 같은 방향을 갖게 된다. 그러나 음성 신호 대 간섭 잡음 비가 낮아질수록 음성 상관 행렬 추정 오차가 커지고 이에 따라 주성분 벡터는 음향 전달 함수와 멀어지게 되며 다채널 위너 필터의 잡음 제거 성능 또한 저하된다.

Reference [1]에서는 주성분 부공간 벡터를 보정하기 위해서 음성 음원의 방향 벡터(steering vector)를 이용하는 방법이 제안되었는데, 주성분 벡터와 방향 벡터가 이루는 각도를 계산하고, 각도가 클수록 주성분 벡터를 방향 벡터 쪽으로 보정하는 정도를 크게 하였다. 이러한 보정 방법은 다채널 위너 필터의 성능을 개선하기는 하나, 주성분 벡터가 왜곡되는 원인에 따라 보정하는 정도를 조절하는 것이 아니라 주성분 벡터가 왜곡된 결과에 따라 보정 정도를 조절하는 방법이다. 본 연구에서는 이러한 방법과 달리 주성분 벡터를 왜곡시키는 원인인 신호 대 간섭 잡음 비와 밀접한 음성 존재 확률을 이용하여 주성분 벡터를 보정하는 정도를 조절하는 방법을 제안하고자 한다. 음성 상관 행렬의 랭크가 1이라는 가정 하에서의 부공간 분해를 통한 다채널 음성 존재 확률을 유도한 후 이를 이용하여 주성분 벡터를 보정하는 방법을 제안한다.

제안하는 방법의 성능을 검증하기 위해 간섭 잡음이 섞인 신호를 다채널 위너 필터에 통과시켜 그 출력 신호의 신호 대 잡음 비와 음성 왜곡과 등의 정량적 지표를 계산하였으며, 그 결과 제안하는 방법을 통해 다채널 위너 필터의 잡음 제거 성능을 향상시킬 수 있음을 확인하였다.

II. 다채널 위너 필터

여러 개의 마이크로부터 취득된 다채널 신호는 다음 식으로 표현될 수 있다.

$$\begin{aligned} Y_i(l,k) &= H_i(l,k)S(k) + N_i(l,k) \\ &= X_i(l,k) + N_i(l,k), \end{aligned} \tag{1}$$

여기서 S는 단일 음성 음원을 나타내고, H_i는 단일 음성 음원으로부터 i번째 마이크까지의 음향 전달 함수를 나타낸다. 그리고 Y_i, X_i, N_i는 i번째 마이크의 신호들을 나타내는데 각각 관찰되는 신호, 음성 성분, 잡음 성분을 나타낸다. l과 k는 각각 시간 프레임과 주파수 대역 인덱스를 나타낸다. M개의 마이크를 이용하여 신호를 수신할 때, Eq. (1)에 표현된 신

호들은 $i=1$에서 M까지 M개의 수식으로 표현되는데 이들을 모아 M차원 벡터로 표현하면 다음 식으로 나타낼 수 있다.

$$\mathbf{y}(l,k)=\mathbf{x}(l,k)+\mathbf{n}(l,k). \tag{2}$$

위 식의 $\mathbf{y,x,n}$은 각각 M차원 벡터를 나타내며 Eqs. (3)~(5)로 나타난다.

$$\mathbf{y}(l,k)=\left[\,Y_1(l,k)\;\;Y_2(l,k)\cdots\;Y_M(l,k)\,\right]^T, \tag{3}$$

$$\mathbf{x}(l,k)=\left[\,X_1(l,k)\;\;X_2(l,k)\cdots X_M(l,k)\,\right]^T, \tag{4}$$

$$\mathbf{n}(l,k)=\left[\,N_1(l,k)\;\;N_2(l,k)\cdots N_M(l,k)\,\right]^T, \tag{5}$$

여기서 $[\,\cdot\,]^T$는 전치 행렬을 나타낸다.

다채널 위너 필터는 여러 마이크 중 특정 마이크에서 관찰되는 신호의 음성 성분을 추정하는 필터로서 추정하는 음성 성분과 실제 참값 사이의 오차 평균을 최소로 하는 수식화 과정을 통해 유도된다. 다채널 위너 필터를 벡터 $\mathbf{w}$로 나타내면 필터를 통과한 출력 신호 $Z(l,k)$는 다음 식으로 나타난다.

$$Z(l,k)=\mathbf{w}^H(l,k)\mathbf{y}(l,k), \tag{6}$$

여기서 $[\,\cdot\,]^H$는 행렬의 원소들을 켤레 복소수로 바꾸고, 행과 열을 전치시키는 연산 즉, Hermitian 행렬로 바꾸는 것을 의미한다. 음성 성분을 추정하는 특정 마이크를 첫 번째 마이크로 정할 경우, 다채널 위너 필터 $\mathbf{w}$는 다음과 같이 나타난다.

$$\begin{aligned}\mathbf{w}(l,k)&=\mathbf{R}_{yy}^{-1}(l,k)\mathbf{R}_{xx}(l,k)\mathbf{e}_1\\&\simeq\mathbf{R}_{yy}^{-1}(l,k)\left(\mathbf{R}_{yy}(l,k)-\mathbf{R}_{nn}(l,k)\right)\mathbf{e}_1,\end{aligned} \tag{7}$$

여기서 $\mathbf{e}_1$은 첫 번째 원소만 1이고 나머지 원소들은 0인 M차원 벡터이고, $\mathbf{R}_{yy},\mathbf{R}_{xx},\mathbf{R}_{nn}$은 다음 식들로 나타나는 상관 행렬들이다. Eq. (7)에서 음성과 잡음의 교차 상관 행렬인 $\mathbf{R}_{xn}$는 음성 상관 행렬에 비해 무시할만한 수준이라고 가정하였다.

$$\mathbf{R}_{yy}(l,k)=E\{\mathbf{y}(l,k)\mathbf{y}^H(l,k)\}, \tag{8}$$

$$\mathbf{R}_{xx}(l,k)=E\{\mathbf{x}(l,k)\mathbf{x}^H(l,k)\}, \tag{9}$$

$$\mathbf{R}_{nn}(l,k)=E\{\mathbf{n}(l,k)\mathbf{n}^H(l,k)\}. \tag{10}$$

Eqs. (8)~(10)에서 $E\{\,\cdot\,\}$는 통계적 기댓값을 나타낸다. 단일 음원으로부터 음성 신호가 전파된다는 가정 하에 신호의 주성분 벡터에서 필터를 구하기 위해 다음 식과 같이 $\mathbf{R}_{yy}$와 $\mathbf{R}_{nn}$을 동시에 대각 행렬로 만드는 동시 대각화를 적용하면 부공간에서 다채널 위너 필터를 구현할 수 있다. 동시 대각화는 일반화된 고유치 분해를 적용함으로써 실행될 수 있다.[11]

$$\mathbf{Q}^H(l,k)\mathbf{R}_{yy}(l,k)\mathbf{Q}(l,k)=\mathbf{\Lambda}_y(l,k), \tag{11}$$

$$\mathbf{Q}^H(l,k)\mathbf{R}_{nn}(l,k)\mathbf{Q}(l,k)=\mathbf{\Lambda}_n(l,k). \tag{12}$$

$$\begin{aligned}\mathbf{R}_{yy}(l,k)&=\overline{\mathbf{Q}}(l,k)\mathbf{\Lambda}_y(l,k)\overline{\mathbf{Q}}^H(l,k)\\&=\sum_{i=1}^{M}\lambda_{y,i}(l,k)\overline{\mathbf{q}}_i(l,k)\overline{\mathbf{q}}_i^H(l,k),\end{aligned} \tag{13}$$

$$\begin{aligned}\mathbf{R}_{nn}(l,k)&=\overline{\mathbf{Q}}(l,k)\mathbf{\Lambda}_n(l,k)\overline{\mathbf{Q}}^H(l,k)\\&=\sum_{i=1}^{M}\lambda_{n,i}(l,k)\overline{\mathbf{q}}_i(l,k)\overline{\mathbf{q}}_i^H(l,k),\end{aligned} \tag{14}$$

여기서 $\mathbf{\Lambda}_y,\mathbf{\Lambda}_n$은 고유치들을 대각 원소로 갖는 대각 행렬이고, $\mathbf{Q}$는 $\mathbf{R}_{yy}$와 $\mathbf{R}_{nn}$을 동시에 대각 행렬로 만드는 행렬이다. $\mathbf{Q}$는 고유 벡터들$(\mathbf{q}_1,\mathbf{q}_2,\cdots,\mathbf{q}_M)$로 이루어진 행렬이며 $\overline{\mathbf{Q}}=\mathbf{Q}^{-H}$의 관계가 성립한다. 신호와 잡음 고유치 비는 $\frac{\lambda_{y,1}}{\lambda_{n,1}}\geq\frac{\lambda_{y,2}}{\lambda_{n,2}}\geq\cdots\geq\frac{\lambda_{y,M}}{\lambda_{n,M}}$ 순으로 정렬되며 여기서 가장 큰 고유치 비율에 해당하는 고유 벡터 즉, $\overline{\mathbf{q}}_1$의 원소들을 아래와 같이 정의한다.

$$\overline{\mathbf{q}}_1(l,k)=\left[\,\bar{q}_{1,1}(l,k)\;\bar{q}_{1,2}(l,k)\;\cdots\bar{q}_{1,M}(l,k)\,\right]^T. \tag{15}$$

Eq. (15)에 나타난 고유 벡터는 음성 신호의 주성분 벡터에 해당하며, 이는 또한 Eq. (1)에 표시된 단일 음

원의 음성이 여러 마이크에 도달하는 음향 전달 함수와 같은 방향을 갖게 된다. Eqs. (11) ~ (15)에서 나타난 상관 행렬 분해를 바탕으로 음성 주성분 벡터 기반의 다채널 위터 필터를 구하면 Eq. (7)은 다음 식으로 표현될 수 있다.

$$\mathbf{w}=\lambda_{n,1}\mathbf{R}_n^{-1}(l,k)\overline{\mathbf{q}}_1(l,k)\left(\frac{\lambda_{y,1}-\lambda_{n,1}}{\lambda_{y,1}}\right)\overline{\mathbf{q}}_1^H(l,k)\mathbf{e}_1. \tag{16}$$

III. 음성 존재 확률을 이용한 주성분 벡터 보정

Eq. (7)에서 음성 상관 행렬을 추정할 때 음성과 간섭 잡음 간에는 상관 관계가 무시할만하다고 가정하였다. 이러한 가정은 음성 신호의 세기가 간섭 잡음 신호보다 월등히 클 경우는 유효하나 간섭 잡음 신호의 세기가 커질 경우는 더 이상 유효하지 않게 되고 이로 인해 음성 상관 행렬 추정 오차가 증가하게 된다. 이러한 경우에 고유치 분해로 계산된 주성분 벡터[Eq. (15)]와 음성 신호의 음향 전달 함수 간의 편차는 커지게 된다.[1]

이러한 문제를 해결하고자 음원의 방향을 알고 있을 때, 음원의 방향 벡터를 이용하여 주성분 벡터를 보정하는 방법을 제안하고자 한다. 방향 벡터는 여러 개의 마이크로 신호를 취득하는 경우, 기준이 되는 마이크(예를 들어, 첫 번째 마이크)에 대한 다른 마이크 신호들의 위상 차이를 나타내는 벡터로서 신호의 공간 정보를 나타내는 벡터이다.[12] 앞에서 언급한 것과 같이 간섭 잡음이 커지게 됨에 따라 주성분 벡터와 방향 벡터 간의 편차가 증가하게 되며, 방향 벡터를 이용하여 주성분 벡터를 보정하고자 한다. 방향 벡터를 이용하여 주성분 벡터를 보정하기 위해서는 보정하는 정도를 결정해주어야 하는데, 보정량을 결정하기 위해서는 신호 대 간섭 잡음 비를 활용할 수 있다. 하지만 신호 대 간섭 잡음 비는 정규화된 수치가 아니라 보정량을 결정하기 어려운 문제가 있다. 따라서 본 논문에서는 신호 대 간섭 잡음 비를 사용하는 대신 음성 존재 확률을 이용하여 주성분 벡터를 보정하는 정도를 결정하는 방법을 제안하고자 한다. 음성 존재 확률은 신호 대 간섭 잡음 비와 밀접한 관계를 가지며, 확률값이므로 0과 1사이 값을 나타내므로 추가적인 정규화 과정없이 보정량을 결정하는데 사용할 수 있다.

음성 존재 확률은 음성이 존재한다는 가설(hypothesis) H_1과 음성이 존재하지 않는다는 가설 H_0에 대해 다음 식과 같이 쓸 수 있다.

$$H_0(l,k):\ \mathbf{y}(l,k)=\mathbf{n}(l,k), \tag{17}$$

$$H_1(l,k):\ \mathbf{y}(l,k)=\mathbf{x}(l,k)+\mathbf{n}(l,k). \tag{18}$$

베이즈 규칙(Bayes' rule)을 사용하면 음성 존재 확률은 다음과 같이 나타낼 수 있다.

$$\begin{aligned}p(l,k)&=\Pr(H_1(l,k)|\mathbf{y}(l,k))\\&=\frac{r(l,k)p(\mathbf{y}(l,k)|H_1(l,k))}{r(l,k)p(\mathbf{y}(l,k)|H_1(l,k))+(1-r(l,k))p(\mathbf{y}(l,k)|H_0(l,k))},\end{aligned} \tag{19}$$

여기서 $r(l,k)$는 사전(*a priori*) 음성 존재 확률($\Pr(H_1(l,k))$)을 나타낸다. 다채널 신호에서 음성과 잡음이 가우시안 분포라고 가정하면 음성 존재 확률은 다음 식으로 나타난다.[7,8]

$$p(l,k)=\left\{1+\frac{1-r(l,k)}{r(l,k)}[1+\xi(l,k)]\exp\left[-\frac{\gamma(l,k)}{1+\xi(l,k)}\right]\right\}^{-1} \tag{20}$$

$$\gamma(l,k)=\mathbf{y}^H(l,k)\mathbf{R}_{nn}^{-1}(l,k)\mathbf{R}_{yy}(l,k)\mathbf{R}_{nn}^{-1}(l,k)\mathbf{y}(l,k)-\mathbf{y}^H(l,k)\mathbf{R}_{nn}^{-1}(l,k)\mathbf{y}(l,k), \tag{21}$$

$$\xi(l,k)=\mathrm{tr}\{\mathbf{R}_{nn}^{-1}(l,k)\mathbf{R}_{yy}(l,k)\}-M. \tag{22}$$

$\xi(l,k)$는 다채널 사전 신호 대 잡음 비(multi-channel *a priori* signal-to-noise ratio)를 나타낸다. Eqs. (11) ~ (14)와 같은 부공간을 고려하면 $\gamma(l,k)$와 $\xi(l,k)$을 다음 식들로 유도할 수 있다.

$$\gamma(l,k) = \mathbf{y}^H(l,k)\mathbf{Q}(l,k) \cdot \Lambda_n^{-1}(l,k)\left(\Lambda_y(l,k)\Lambda_n^{-1}(l,k) - \mathbf{I}_M\right)\mathbf{Q}^H(l,k)\mathbf{y}(l,k), \tag{23}$$

$$\xi(l,k) = \mathrm{tr}\left(\mathbf{Q}(l,k)\boldsymbol{\Lambda}_n^{-1}(l,k)\boldsymbol{\Lambda}_y(l,k)\mathbf{Q}^H(l,k)\right) - M. \tag{24}$$

단일 음원으로부터 음성이 마이크들에 도달할 경우 음성 상관 행렬 $\mathbf{R}_{xx}(l,k)$의 랭크는 이론적으로 1이 되고, 이 경우 $\mathbf{Q}(l,k)$의 고유 벡터 중 주성분 고유벡터($\mathbf{q}_1(l,k)$)만 고려하면 Eqs. (23)과 (24)로 나타난 $\gamma(l,k)$와 $\xi(l,k)$는 아래와 같은 간단한 식으로 유도될 수 있다.

$$\gamma(l,k) = \frac{\lambda_{y,1}(l,k) - \lambda_{n,1}(l,k)}{\lambda_{n,1}^2(l,k)}\left(\mathbf{y}^H(l,k)\mathbf{q}_1(l,k)\right)^2, \tag{25}$$

$$\xi(l,k) = \frac{\lambda_{y,1}(l,k)}{\lambda_{n,1}(l,k)} \parallel \mathbf{q}_1(l,k) \parallel^2. \tag{26}$$

방향 벡터를 이용하여 Eq. (15)에서 구한 주성분 벡터 $\overline{\mathbf{q}}_1(l,k)$를 보정하기 위해서는 $\overline{\mathbf{q}}_1(l,k)$와 방향 벡터 $\mathbf{v}(k)$를 변형하는 과정이 필요하다. 먼저 $\overline{\mathbf{q}}_1(l,k)$을 $\mathbf{v}(k)$와 같이 위상 정보만 갖도록 하기 위해서 다음 식과 같이 $\overline{\mathbf{q}}_1(l,k)$ 벡터의 각 원소들을 각각의 절대값으로 나누어준다.

$$\overline{\overline{\mathbf{q}}}(l,k) = \left[\frac{\bar{q}_1(l,k)}{|\bar{q}_1(l,k)|} \quad \frac{\bar{q}_2(l,k)}{|\bar{q}_2(l,k)|} \quad \cdots \quad \frac{\bar{q}_M(l,k)}{|\bar{q}_M(l,k)|}\right]^T. \tag{27}$$

방향 벡터의 첫 번째 원소 위상을 Eq. (27)의 위상과 동일하게 만들기 위해 다음 식과 같이 방향 벡터의 각 원소에 Eq. (27)에 나타난 벡터의 첫 번째 원소를 곱해준다.

$$\mathbf{v}'(l,k) = \left[\overline{\bar{q}}_1(l,k)v_1(l,k) \quad \cdots \overline{\bar{q}}_1(l,k)v_M(l,k)\right]^T. \tag{28}$$

$\overline{\bar{q}}_1(l,k)$는 $\overline{\overline{\mathbf{q}}}(l,k)$의 첫 번째 원소이다.

Eqs. (20)과 (25), (26)으로 구한 음성 존재 확률과 Eqs. (27)과 (28)을 이용하여 다음 식과 같이 주성분 벡터를 보정할 수 있다.

$$\overline{\overline{\mathbf{q}}}'(l,k) = p(l,k)\overline{\overline{\mathbf{q}}}(l,k) + (1-p(l,k))\mathbf{v}'(l,k). \tag{29}$$

위 식에 나타난 것과 같이 음성 존재 확률이 1에 가까울 때는 보정이 거의 이루어지지 않게 되고, 음성 존재 확률이 0에 가까운 경우는 보정된 벡터는 방향 벡터에 가깝게 된다. 방향 벡터를 이용하여 주성분 벡터의 위상만 수정하고자 하므로 보정된 벡터 각 원소들의 절대값이 Eq. (15) 원소들의 절대값이 되도록 스케일링한다.

IV. 실험결과

실험을 위한 다채널 신호를 생성하기 위해 RWCP sound scene database에 포함되어 있는 임펄스 응답을 사용하였다.[13] 5.66 cm의 등간격을 이루는 7개의 선형 어레이 마이크의 정면에서 2 m 떨어진 곳에 음성 음원이 배치되어 있고 간섭 잡음은 음성의 방향과 40°를 이루는 방향에서 음성과 마찬가지로 2 m 떨어져서 입사하는 상황을 가정하였다. 샘플링 주파수는 16 kHz이고, 반향 시간은 300 ms이다. 이러한 임펄스 응답을 음성 신호 및 간섭 잡음 신호와 콘볼루션하고 신호 대 간섭 잡음 비(Signal to Interference Noise Ratio, SINR)에 따라 혼합하여 다채널 신호를 재현하였다. 간섭 잡음으로는 백색 잡음과 음성 간섭 잡음(competing speech)을 사용하였으며, 0 dB, 5 dB, 10 dB의 SINR이 되도록 다채널 신호를 생성하여 실험하였다.

잡음 제거 성능을 측정하기 위해 SINR gain, Mel-Frequency Cepstral Coefficient(MFCC) 거리를 사용하였다. 사전 음성 존재 확률 $r(l,k)$는 Reference [8]에서와 같이 0.4로 두고 실험하였다. 사전 음성 존재 확률 값에 따라 음성 존재 확률이 영향을 받으나 Eq. (20)에서 확인할 수 있듯이 $\frac{\gamma(l,k)}{1+\xi(l,k)}$가 주된 영향을 미치게 되므로 사전 음성 존재 확률을 0이나 1에 가깝

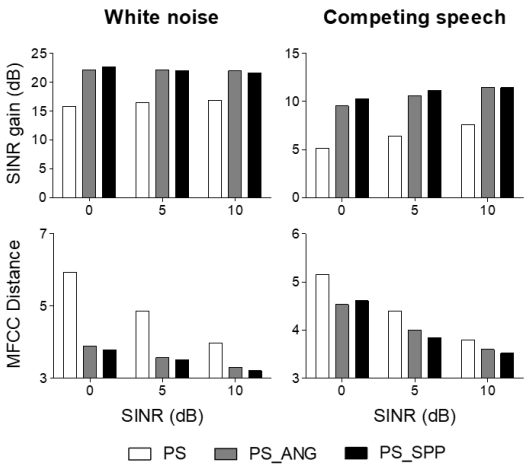

Fig. 1. Performance evaluation using SINR gain and MFCC distance for white noise and competing speech.

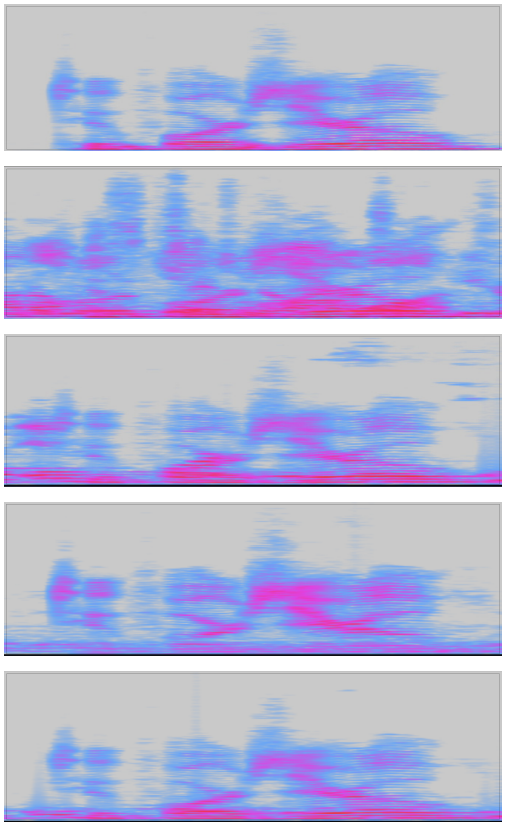

Fig. 2. (Color available online) Spectrograms signals (From top): clean target speech, corrupted speech followed by noise suppressed signals by PS, PS_ANG, PS_SPP, respectively.

게 설정하지 않는 한 성능에 미치는 영향은 크지 않다. 실험 결과는 Fig. 1에 그래프로 표시하였다. Fig. 1에 PS로 표시된 것은 주성분 벡터를 이용한 다채널 위너 필터이고, PS_ANG는 주성분 벡터와 방향 벡터의 각도를 이용하여 주성분 벡터를 보정한 결과이며, PS_SPP는 본 논문에서 제안하는 음성 존재 확률을 이용하여 주성분 벡터를 보정한 다채널 위너 필터의 결과이다.

백색 잡음과 음성 간섭 잡음 두 경우 모두 주성분 다채널 위너 필터(Principal Subspace, PS)에 비해 주성분을 보정한 필터의 성능이 SINR gain과 MFCC 거리 측면에서 우수한 것으로 나타났다. 제안하는 보정 방법인 PS_SPP는 MFCC 거리 측면에서 음성 간섭 잡음 0 dB인 경우에는 PS_ANG보다 약간 성능이 떨어지는 것으로 나타나는데 이것은 낮은 SINR 환경 하에서 음성 존재 확률 추정이 저하된 결과로 추정할 수 있다. 나머지 경우는 제안하는 PS_SPP가 PS_ANG보다 우수함을 알 수 있다. Fig. 2에 음성 간섭 잡음, 0 dB SINR인 경우의 스펙트로그램을 나타내었다. 가로축은 시간을 나타내고 세로축은 0 kHz부터 8 kHz까지의 주파수를 나타낸다.

V. 결 론

다채널 위너 필터는 신호의 부공간 분해를 통해 구현할 수 있는데 주성분 부공간에서 음성 신호를 추정하여 효과적으로 잡음을 제거할 수 있다. 이론상으로 음성 주성분 벡터는 음향 전달 함수와 같은 위상을 나타내야 하지만 SINR이 낮은 시간-주파수 영역에서는 오차로 인한 성능 저하가 발생하며 이를 해결하기 위해 음성의 방향 벡터와 음성 존재 확률을 이용하여 음성 주성분 벡터를 보정하고자 하였다. 제안하는 방법을 적용하여 실험한 결과 SINR gain, MFCC거리 측면에서 잡음 제거 성능을 향상시킬 수 있음을 확인하였다.

References

1. G. Kim, “Interference suppression using principal subspace modification in multichannel Wiener filter and its application to speech recognition,” ETRI J. **32**, 921-931 (2010).
2. K. Ngo, A. Spriet, M. Moonen, J. Wouter, and S. H. Jensen, “Incorporating the conditional speech presence

probability in multi-channel Wiener filter based noise reduction in hearing aids," Eurasip J. Advances in Signal Processing, **2009**, 1-11 (2009).
3. K. Ngo, M. Moonen, S. H. Jensen, and J. Wouters, "A flexible speech distortion weighted multi-channel Wiener filter for noise reduction in hearing aids," ICASSP. 2528-2531 (2011).
4. T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Generalized multichannel wiener filter for spatially distributed microphones," Speech Communication;. 11. ITG Symposium, 1-4 (2014).
5. 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 Trans. Audio, Speech, and Lang. Process. **22**, 785-798 (2014).
6. Z. Wang, E. Vincent, R. Serizel, and Y. Yan, "Rank-1 constrained multichannel Wiener filter for speech recognition in noisy environments," Computer Speech & Language, **49**, 37-51 (2018).
7. S. Bagheri and D. Giacobello, "Exploiting multi-chaannel speech presence probability in parametric multi-channel Wiener filter," Interspeech, 101-105 (2019).
8. M. Souden, J. Chen, J. Benesty, and S. Affes, "Gaussian model-based multichannel speech presence probability," IEEE Trans. Audio, Speech, Lang. Process. **18**, 1072-1077 (2010).
9. Y. G. Jin, J. W. Shin, and N. S. Kim, "Decision-directed speech power spectral density matrix estimation for multichannel speech enhancement," J. Acoust. Soc. Am. **141**, EL234 (2017).
10. S. Jeong and Y. Kim, "An optimally-modified multi-channel Wiener filter using speech presence probability" (in Korean), Smart Media J. **7**, 9-15 (2018).
11. G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd ed.* (Johns Hopkins University Press, Baltimore, 1996), Chap. 8.
12. H. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory* (Wiley, Hoboken, 2002), Chap. 2.
13. S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," Proc. the 2nd Int. Conf. LREC. 965-968 (2000).

저자 약력

▸ 김 기 백 (Gibak Kim)

1994년 2월 : 서울대학교 전자공학과 학사
1996년 2월 : 서울대학교 전자공학과 석사
2007년 8월 : 서울대학교 전기컴퓨터공학부 박사
2011년 9월 ~ 현재 : 숭실대학교 전기공학부 교수