

정규논문 (Regular Paper)

방송공학회논문지 제18권 제3호, 2013년 5월 (JBE Vol. 18, No. 3, May 2013)

<http://dx.doi.org/10.5909/JBE.2013.18.3.455>

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

공간 필터와 결합된 음성 왜곡 가중 다채널 위너 필터에서의 신호 대 잡음 비에 의한 가중치 결정 방법

김 기 백^{a)‡}

SNR-based Weight Control for the Spatially Preprocessed Speech Distortion Weighted Multi-channel Wiener Filtering

Gibak Kim^{a)‡}

요 약

본 논문에서는 여러 개의 마이크를 이용하여 잡음을 제거하는 방법인 공간 필터로 전처리된 신호를 입력으로 하는 음성 왜곡 가중 다채널 위너 필터 (Spatially Preprocessed Speech Distortion Weighted Multi-channel Wiener Filter: SP-SDW-MWF)에 대해 소개하고, 가중치를 결정하는 방법을 제안한다. SP-SDW-MWF는 마이크로폰 어레이를 이용한 잡음 제거 알고리즘으로서 마이크로폰 불일치와 같은 오차에 강인한 것으로 알려져 있다. SP-SDW-MWF는 필터 계수를 최적화할 때 음성 왜곡과 잡음 제거에 대한 기준으로 나누어 가중치를 두고 있다. 이러한 가중치를 결정하기 위해, 본 논문에서는 전력 스펙트럼 밀도 오차를 평가 척도로 사용하여 마이크로폰으로부터 입력된 음성 신호와 잡음의 전력 스펙트럼 밀도의 비 (a priori SNR)를 이용하는 방법을 제안한다. 실험결과에서 나타난 바와 같이 a priori SNR에 따라 가변적인 가중치를 사용하는 것이 고정된 값을 가중치로 사용하는 것보다 향상된 성능을 보임을 알 수 있다.

Abstract

This paper introduces the Spatially Preprocessed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF) for multi-microphone noise reduction and proposes a method to determine the speech distortion weights. The SP-SDW-MWF is known as a robust noise reduction algorithm against the error caused by the mismatch in microphones. The SP-SDW-MWF adopts weights which determine the amount of noise reduction at the expense of introducing speech distortion in the noise-suppressed speech. In this paper, we use the error of power spectral density between the estimated signal and the desired signal as the evaluation measure. Thus the a priori SNR is used to control the speech distortion weights in the frequency domain. In the experimental results, the proposed method yields better result in terms of MFCC distortion compared to the conventional method.

Keyword : Noise reduction, multi-channel Wiener filter, speech distortion weights

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

※ 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행되었습니다 (2013-027840).

· Manuscript received January 17, 2013 Revised April 15, 2013 Accepted April 15, 2013

1. 서론

잡음 제거는 실제 환경에서 음성 통신이나 음성 인식을 위하여 필요한 요소 중의 하나이다. 실제 환경에서는 여러 가지 잡음의 영향을 많이 받게 되고, 잡음 환경에 따라서는 심각한 성능 저하를 초래하는 경우도 많다. 이러한 잡음 제거의 중요성으로 인해 다양한 잡음 제거 알고리즘이 연구되어 왔다^[1-6]. 하드웨어의 비약적인 발전에 힘입어 최근에는 마이크로폰 어레이를 이용하여 음성 왜곡을 최소화하면서 높은 잡음 제거 성능을 보이는 알고리즘들이 많이 연구되고 있다. 하나의 마이크로폰을 사용하는 경우는 음성과 잡음의 시간-주파수 영역의 정보만 이용하는 반면, 여러 개의 마이크로폰을 이용하는 경우는 신호의 공간적인 정보를 이용하여 필터링하므로 음성 왜곡을 최소화하면서도 높은 잡음 제거 성능을 얻을 수 있게 된다. 마이크로폰 어레이를 이용하는 잡음 제거 알고리즘으로는 고정 빔포밍 (Fixed beamforming), 적응 빔포밍 (Adaptive beamforming), 포스트 필터링 (Post-filtering), 그리고 다채널 위너 필터링 (Multi-channel Wiener filtering) 등이 있다^[7]. 최근 몇 년 동안 다채널 위너 필터링에 관한 연구가 많이 이루어졌는데 GSVD (Generalized Singular Value Decomposition), QR 분해 등을 이용한 구현 방법들이 제안되었다^[8,9]. 이 방법들은 빔포밍 알고리즘과는 달리 음성 신호의 방향 정보를 사용하지 않고 음성 신호 성분과 잡음 성분의 2차 통계 특성을 추정하여 필터를 구하기 때문에 방향 감지 오류 등에 의한 성능 저하를 유발하지 않는다는 장점을 갖고 있다.

최근에 제안된 SP-SDW-MWF는 어레이 마이크로폰 신호를 필터의 입력으로 바로 사용하지 않고 음성 신호의 방향 정보를 이용하여 공간 필터링을 거친 신호를 필터의 입력으로 사용한다^[10]. 또한 이 방법은 음성 성분의 왜곡과 잡음 제거 정도를 가중치를 두어 조절할 수 있다. 목적 신호와 추정 신호 간의 최소 자승 오차 (Mean Squared Error: MSE)가 성능 판단의 기준이 될 때는 이러한 가중치가 필요없지만, 음성의 명료도 (intelligibility)의 향상, SNR 향상, 전력 스펙트럼 밀도 오차의 최소화 등의 다른 여러 가지 성능을 측정할 때는 적절한 음성 왜곡 가중치를 사용하여

성능 향상을 도모할 수 있다. 가중치는 평가 척도에 따라 고정된 가중치를 사용할 수도 있으나, 시간, 주파수에 따라 가변적인 가중치를 사용할 수도 있다.

Spriet 등에 의해 처음 제안되었던 SP-SDW-MWF에서는 가중치를 도입한다는 내용이 포함되어 있지만 가중치를 어떻게 정하느냐에 관한 문제에 대해서는 다루지 않았다. 잡음 제거 성능을 더 중요시하는 응용 분야에서는 음성 왜곡이 어느 정도 발생하더라도 잔류 잡음을 최소화하도록 가중치를 결정하여야 하겠고, 음성 인식과 응용 분야와 같이 잡음이 어느 정도 있더라도 원래 음성이 갖고 있는 정보를 잃어버려서는 안되는 경우에는 음성 왜곡을 최소화하도록 가중치를 정해야 할 것이다. 그러나 어느 경우에서나 잡음 환경에 따라 최적의 가중치가 달라져서 자동으로 가중치를 정해주는 것은 어려운 문제라고 할 수 있다.

본 논문에서는 목적 신호와 추정 신호간의 전력 스펙트럼의 차이를 유사도 측정의 기준으로 하는 경우 음성 왜곡 가중치를 결정하는 방법을 제안하고자 한다. 전력 스펙트럼은 음성 신호의 특성을 나타내는 중요한 척도가 되며, 기존의 단일 마이크 잡음 제거 필터 추정에서도 전력 스펙트럼 오차를 최소화하는 방법이 널리 사용되었다^[3,4]. 잡음 제거 시스템에서 출력 신호 (추정신호)와 목적 신호의 전력 스펙트럼 밀도 오차로 성능을 측정할 경우는 모든 주파수 대역에 대해 고정된 가중치를 두는 것보다 각 주파수 대역별 a priori SNR (사전 신호 대 잡음 비)에 따라 가중치를 다르게 하여 성능 향상을 기대할 수 있다. a priori SNR은 SP-SDW-MWF에서 전처리로 사용되는 GSC (Generalized Sidelobe Canceller)^[11] 구조에서 고정 빔포머와 차단 행렬 (blocking matrix)의 출력을 이용하여 추정할 수 있다. 이렇게 각 주파수 대역에서 추정된 a priori SNR에 따라 음성 왜곡 가중치를 결정할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 SP-SDW-MWF를 공간상에서의 전처리 (GSC)와 음성 왜곡 가중 다채널 위너 필터 (SDW-MWF)로 나누어 살펴본다. 3장에서는 전력 스펙트럼 밀도 오차를 성능 평가의 기준으로 할 때의 가중치와 a priori SNR과의 관계를 유도하고, 고정 빔포머와 차단 행렬의 출력으로부터 a priori SNR을 추정하는 방법에 대해 설명한다. 4장에서는 실험 결과에 대해 설

명하고, 5장에서 논문의 결론을 맺는다.

II. SP-SDW-MWF

SP-SDW-MWF는 그림 1과 같이 GSC와 SDW-MWF의 결합으로 나타낼 수 있다. 본 장에서는 GSC와 SDW-MWF에 대해 각각 알아보고, 이 두 알고리즘의 결합인 SP-SDW-MWF에 대해 설명하고자 한다.

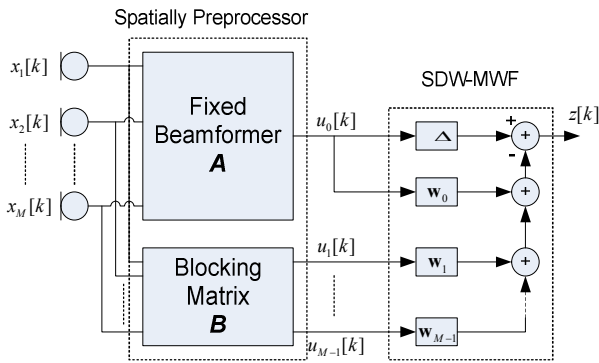


그림 1. 공간 필터로 전처리된 신호를 입력으로 하는 음성 왜곡 가중 다채널 위너 필터
 Fig. 1. Spatially preprocessed speech distortion weighted multi-channel Wiener filter (SP-SDW-MWF)

1. GSC

GSC는 고정 빔포머로부터 음성 신호의 레퍼런스를 만들고, 이 신호를 입력으로 하는 적응 필터를 이용하여 잡음 성분을 제거하는 시스템이다^[11]. 먼저 M개의 마이크로폰 신호에 대해 적절한 시간 지연을 줌으로써 모든 마이크로폰들이 음성 신호가 위치한 방향으로 향하게 한다. 즉, 음성 신호 성분에 대해서는 모든 마이크로폰 신호의 위상이 동일하게 되도록 만들어 준다. 그런 다음 고정 빔포머를 통과하여 적응 필터의 레퍼런스 입력 신호로 사용한다. 여기서 사용될 고정 빔포머는 마이크로폰 불일치와 같은 신호 모델 오류에서도 음성 신호 성분의 왜곡이 많지 않아야 한다. Delay-and-sum 빔포머는 간단하면서도 작은 개수의 마이크를 사용하는 마이크로폰 어레이에서도 강인한 성능을 가

지므로 본 시스템의 고정 빔포머로 사용한다. 차단 행렬 (Blocking Matrix)은 M개의 마이크로폰 신호를 입력으로 받아들여 목적 음성 신호 성분은 차단하고 잡음 성분만 통과시켜, 잡음 제거 적응 필터의 입력으로 사용될 M-1개의 잡음 레퍼런스 신호를 추출한다. 실제의 경우에는 차단 행렬을 통과한 잡음 레퍼런스 신호에 음성 신호가 섞여 있는데 이를 음성 누설 (speech leakage)라고 한다. 음성 누설이 잡음 제거 적응 필터의 입력으로 들어가면 적응 필터의 출력 신호에 음성 왜곡을 발생시킨다. 이러한 음성 왜곡을 방지하기 위해서 음성 구간 검출기를 이용하여 음성 신호가 없는 구간에서만 잡음 제거 적응 필터의 계수를 변화시키거나, 누설 음성 신호에 비해 잡음 성분이 월등히 크도록 차단행렬을 설계하는 등의 여러 방법들이 제안되어 왔다 [12].

2. 음성 왜곡 가중 다채널 위너 필터 (SDW-MWF)

최근에 마이크로폰 어레이를 이용한 잡음 제거 알고리즘으로 최소 평균 자승 오차 (Minimum Mean Squared Error: MMSE) 추정에 근거한 다채널 위너 필터 (Multi-channel Wiener Filter: MWF)가 제안되었다^[8]. 이 방법은 빔포밍과는 달리 화자의 위치와 같은 사전 정보를 필요로 하지 않기 때문에 방향 검지 오류가 미치는 성능 저하를 겪지 않는다는 장점을 갖고 있다. M개의 마이크로폰을 사용할 때 필터의 입력 신호 \mathbf{u} 와 L차 필터 계수 \mathbf{w} 는 다음과 같이 정의한다.

$$\mathbf{w}^H = [\mathbf{w}_1^H \mathbf{w}_2^H \dots \mathbf{w}_M^H] \quad (1)$$

$$\mathbf{w}_i = [w_i[0] w_i[1] \dots w_i[L-1]]^T \quad (2)$$

$$\mathbf{u}^H[k] = [u_1^H[k] u_2^H[k] \dots u_M^H[k]] \quad (3)$$

$$\mathbf{u}_i[k] = [u_i[k] u_i[k-1] \dots u_i[k-L+1]]^T \quad (4)$$

$$u_i[k] = u_i^s[k] + u_i^n[k], \quad i = 1, \dots, M. \quad (5)$$

여기서 위첨자 H는 행렬의 Hermitian 연산을 의미한다.

즉, 전치행렬을 구하고 모든 원소들을 켈레의 복소수로 바꾸는 것을 말한다. 식 (5)와 같이 입력 신호(u_i)는 음성 성분(u_i^s)과 잡음 성분(u_i^n)의 합으로 볼 수 있다. 추정하고자 하는 목적 신호가 첫 번째 마이크로폰 신호에 어떤 시간 지연이 적용된 신호라고 하면, MWF를 구하기 위한 평균 자승 오차 (MSE)는 다음과 같이 표현될 수 있다.

$$E\{e[k]^2\} = E\{[u_i^s[k-\Delta] - \mathbf{w}^H \mathbf{u}[k]]^2\} \quad (6)$$

MSE를 최소화하는 최적의 필터 계수는 다음과 같다.

$$\mathbf{w}_{MWF} = E\{\mathbf{u}[k]\mathbf{u}^H[k]\}^{-1} E\{\mathbf{u}[k]u_i^{s,*}[k-\Delta]\}. \quad (7)$$

여기서 평균 자승 오차는 다음과 같이 음성 왜곡 에너지와 잔류 잡음 에너지로 분리할 수 있다.

$$E\{e[k]^2\} = \epsilon_s^2 + \epsilon_n^2 \quad (8)$$

$$\epsilon_s^2 = E\{[u_i^s[k-\Delta] - \mathbf{w}^H \mathbf{u}^s[k]]^2\}, \epsilon_n^2 = E\{[\mathbf{w}^H \mathbf{u}^n[k]]^2\} \quad (9)$$

음성 왜곡을 줄이는 것과 잡음을 줄이는 것은 서로 trade-off 관계를 갖는다. 즉, 잡음을 많이 줄이고자 하면 음성 왜곡이 발생하고, 음성 왜곡을 줄이고자 하면 잔류 잡음이 크게 나타난다. 따라서 가중치를 부여하여 음성 왜곡과 잔류 잡음을 조절할 수 있는 필터를 다음과 같이 구할 수 있다.

여기서 위첨자 *는 켈레 복소수를 의미한다. 이러한 필터를 음성 왜곡 가중 MWF (Speech Distortion Weighted MWF: SDW-MWF) 라 한다. 가중치 μ 는 0보다 크거나 같은 실수값을 갖는데, 가중치가 클수록 잔류 잡음이 작아지면서 음성 왜곡이 많이 발생하고, 가중치가 작을수록 음성 왜곡이 작아지는 대신 잔류 잡음이 커진다. 즉, 가중치가

무한대가 되면 음성 왜곡은 무시되고 모든 노력이 잡음 제거에 집중되어 출력값이 0이 된다. 반대로 가중치가 0이 되면 음성 왜곡이 최소가 되도록 하기 때문에 입력값이 그대로 출력으로 나온다. 가중치가 1일 때는 처음에 언급한 MMSE 방법이 된다.

3. SP-SDW-MWF

앞 절에서 언급한 SDW-MWF는 어레이 마이크로폰 신호를 입력으로 하고 음성 신호와 잡음의 2차 통계적 특성을 추정하여 필터 계수를 추정한다. 음성 신호의 방향에 대한 정보가 주어진다면, 공간적 필터링을 거치고 난 신호를 MWF의 입력으로 사용하여 성능 향상을 도모할 수 있다. 공간 필터링의 전처리 과정은 1절에서 언급한 GSC처럼 고정 빔포머와 차단 행렬을 사용하여 1개의 음성 레퍼런스 신호와 M-1개의 잡음 레퍼런스 신호를 얻게 되고, 이를 MWF의 입력으로 사용하게 된다. MWF는 M개의 입력 신호를 받아들여 음성 레퍼런스의 잡음 성분을 추정하고 음성 레퍼런스로부터 추정된 값을 차감하여 출력을 얻는다. M개의 입력 신호에 대해 음성 성분과 잡음 성분의 상관 행렬 (correlation matrix)을 추정하여 구한 최적의 필터 계수는 다음과 같다.

III. 음성 왜곡 가중치의 결정 방법

본 장에서는 SP-SDW-MWF 필터 계수를 주파수 영역에서 나타낸다. 코히어런스 행렬을 이용하여 상관 행렬로부터 전력 스펙트럼 밀도를 분리한다.

1. 주파수 영역에서의 필터 계수

SDW-MWF는 식 (8), 식 (9)와 같이 평균 자승 오차를

$$\mathbf{w} = E\{\mathbf{u}^s[k]\mathbf{u}^{s,H}[k] + \mu\mathbf{u}^n[k]\mathbf{u}^{n,H}[k]\}^{-1} E\{\mathbf{u}^s[k]u_i^{s,*}[k-\Delta]\}. \quad (10)$$

$$\mathbf{w} = (E\{\mathbf{u}^s[k]\mathbf{u}^{s,H}[k]\}/\mu + E\{\mathbf{u}^n[k]\mathbf{u}^{n,H}[k]\})^{-1} E\{\mathbf{u}^n[k]u_0^{n,*}[k-\Delta]\}. \quad (11)$$

$$\mathbf{W}(f) = [\epsilon\{\mathbf{U}^s(f)\mathbf{U}^{s,H}(f)\}/\mu + \epsilon\{\mathbf{U}^n(f)\mathbf{U}^{n,H}(f)\}]^{-1} \epsilon\{\mathbf{U}^n(f)\mathbf{U}^{n,H}(f)\} \mathbf{e}_1. \quad (12)$$

음성 왜곡 에너지와 잔류 잡음 에너지의 두 기준으로 나누고 적당한 가중치를 두어 최적화하는 것이다. SP-SDW-MWF로 구한 필터를 주파수 영역에서 나타내면 다음 식과 같다.

여기서 $\mathbf{W}(f) = [W_1(f) \ W_2(f) \ \dots \ W_M(f)]^T$ 이고, $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]^T$ 이다. 균일한 (homogeneous) 공간이라고 가정하면 모든 채널 신호의 전력 스펙트럼 밀도는 동일하며 식 (12)의 음성 성분과 잡음 성분의 상관 행렬들은 다음과 같이 표현할 수 있다. 이 후로는 식을 간단히 표현하기 위해 주파수 성분을 나타내는 기호인 f 는 생략하기로 한다.

$$\epsilon\{\mathbf{U}^s\mathbf{U}^{s,H}\} = P_X^s \begin{bmatrix} \mathbf{A}^H \mathbf{\Gamma}^s \mathbf{A} \ \mathbf{A}^H \mathbf{\Gamma}^s \mathbf{B} \\ \mathbf{B}^H \mathbf{\Gamma}^s \mathbf{A} \ \mathbf{A}^H \mathbf{\Gamma}^s \mathbf{B} \end{bmatrix} \quad (13)$$

$$\epsilon\{\mathbf{U}^n\mathbf{U}^{n,H}\} = P_X^n \begin{bmatrix} \mathbf{A}^H \mathbf{\Gamma}^n \mathbf{A} \ \mathbf{A}^H \mathbf{\Gamma}^n \mathbf{B} \\ \mathbf{B}^H \mathbf{\Gamma}^n \mathbf{A} \ \mathbf{A}^H \mathbf{\Gamma}^n \mathbf{B} \end{bmatrix}. \quad (14)$$

여기서 \mathbf{A} 와 \mathbf{B} 는 각각 고정 빔포머와 차단 행렬을 나타내고, P_X^s 와 P_X^n 은 음성 신호와 잡음 신호의 전력 스펙트럼 밀도이며, $\mathbf{\Gamma}^s$ 와 $\mathbf{\Gamma}^n$ 는 어레이 마이크로폰에 입력되는 음성 신호와 잡음 신호의 코히어런스 행렬 (Coherence Matrix)이다. 코히어런스 행렬의 원소들은 두 마이크로폰 사이의 복소 코히어런스 함수 (Complex Coherence Function)이며 k 번째 마이크와 l 번째 마이크의 복소 코히어런스 함수는 다음과 같이 정의할 수 있다.

$$\Gamma_{kl} = \frac{P_{X_k X_l}}{P_{X_k} P_{X_l}}. \quad (15)$$

여기서 $P_{X_k X_l}$ 는 상호 전력 스펙트럼 밀도이다. 음성 신호의 음원 전력 스펙트럼을 P_{S_s} , 음원으로부터 마이크로폰 어레이까지의 M 채널 전달함수를 \mathbf{H} 라 하면 그림 1의 필터 계수는 다음과 같다.

$$\mathbf{W} = \mathbf{e}_1 - \bar{\mathbf{W}} \quad (16)$$

$$\bar{\mathbf{W}} = \Lambda (P_X^n \mathbf{\Gamma}_U^n)^{-1} \mathbf{H}_U \frac{P_{S_d}}{P_{S_d} + \mu \Lambda} \mathbf{H}_{U,1}^* \quad (17)$$

$$\mathbf{H}_U \equiv \begin{bmatrix} \mathbf{A}^H \\ \mathbf{B}^H \end{bmatrix} \mathbf{H} \quad (18)$$

$$\mathbf{\Gamma}_U^n \equiv \begin{bmatrix} \mathbf{A}^H \\ \mathbf{B}^H \end{bmatrix} \mathbf{\Gamma}^n [\mathbf{A} \ \mathbf{B}] \quad (19)$$

$$\Lambda \equiv (\mathbf{H}_U^H (P_X^n \mathbf{\Gamma}_U^n)^{-1} \mathbf{H}_U)^{-1} \quad (20)$$

$$\mathbf{H}_{U,1}^* = \mathbf{H}_U^H \mathbf{e}_1. \quad (21)$$

2. 음성 왜곡 가중치와 a priori SNR과의 관계

SP-SDW-MWF의 출력 신호와 목적 신호간의 평균 자승 오차 (mean square error)를 최소화하는 가중치의 값은 1이다. 그러나 음성 신호처리에서는 신호간의 평균 자승 오차보다도 전력 스펙트럼의 오차가 더 중요한 경우가 많다. 음성 신호처리에서 많은 경우에 유사도 측정을 위해서 캡스트럴 (cepstral)영역의 특징 벡터를 사용하는데, 캡스트럴 영역에서의 L2-거리는 신호의 로그 전력 스펙트럼의 거리와 같음을 알 수 있다^[13].

$$d_2^2 = \int_{-\pi}^{\pi} |\log P_s(\omega) - \log P_{s'}(\omega)|^2 \frac{d\omega}{2\pi} = \sum_{n=-\infty}^{\infty} (c_n - c'_n)^2 \quad (22)$$

$P_s(\omega)$ 와 $P_{s'}(\omega)$ 는 신호의 전력 스펙트럼을 나타낸 것이고, c_n 과 c'_n 은 캡스트럴 계수를 나타낸 것이다.

SP-SDW-MWF의 출력 신호와 원하는 신호의 전력 스펙트럼 오차를 유도하여 가중치와 a priori SNR과의 관계를 알아보기로 한다. 시스템의 출력 신호는 다음과 같이 입력 신호와 필터 계수로 표현할 수 있다.

$$\mathbf{Z} = U_0 - \mathbf{W}^H \mathbf{U} = \bar{\mathbf{W}}^H \mathbf{U} \quad (23)$$

출력 신호의 전력 스펙트럼 밀도는 다음 식과 같이 구할

수 있다.

$$P_Z = \overline{\mathbf{W}}^H \epsilon \{ \mathbf{U} \mathbf{U}^H \} \overline{\mathbf{W}}. \quad (24)$$

출력 신호와 목적 신호의 전력 스펙트럼 밀도의 비를 구해보면 다음과 같다.

$$\frac{P_Z}{P_X} = \frac{\gamma + \Lambda'}{(\gamma + \mu \Lambda')^2} \quad (25)$$

$$\Lambda' = (\mathbf{H}^H (\mathbf{I})^{-1} \mathbf{H})^{-1} \quad (26)$$

$$\gamma = \frac{P_X^s}{P_X^n}. \quad (27)$$

M채널 전달함수 H는 전력 스펙트럼 밀도의 손실이 없다고 가정하였다. 식(25)에서 보듯이 전력 스펙트럼의 오차는 a priori SNR을 나타내는 γ , 음성 왜곡 가중치를 나타내는 μ , 음성 신호의 공간 전달 함수를 나타내는 H, 그리고 잡음의 공간적 특성을 나타내는 코히어런스 행렬 \mathbf{I}^n 에 의해 결정된다. 식(26)에 나타난 Λ' 는 그림 1에서 GSC에 의한 잡음 제거량을 나타낸다. γ 가 작을 때, 즉 잡음이 많이 섞여 있을 때는 μ 를 크게 하여 잡음 제거에 더 많은 가중치를 두고, γ 가 클 때, 즉 잡음이 별로 없을 때는 μ 를 작게 하여 음성 왜곡이 최소화 되도록 하여야 전력 스펙트럼 밀도의 왜곡을 효과적으로 줄일 수 있게 된다. 공간상에서 상관성이 없는 잡음인 경우는 코히어런스 행렬은 단위 행렬이 되고, $|\mathbf{H}^H \mathbf{H}| = M$ 이라 하면 $\Lambda' = 1/M$ 이 된다. 이 때 전력 스펙트럼 오차인 식(25)는 다음과 같이 된다.

$$\frac{P_Z}{P_X} = \frac{\gamma + 1/M}{(\gamma + \mu/M)^2}. \quad (28)$$

위의 식으로부터 최적의 가중치 μ 를 다음과 같이 구할 수 있다.

$$\mu = \frac{\sqrt{\gamma^2 + \gamma \Lambda'} - \gamma}{\Lambda'}. \quad (29)$$

3. a priori SNR 추정값을 이용한 기준치 조절

제안하는 알고리즘에서는 GSC 고정 빔포머의 출력과 차단 행렬의 출력을 이용하여 마이크로폰 신호의 a priori SNR을 추정할 수 있다. 고정 빔포머의 출력 신호와 차단 행렬의 출력 신호에 대한 전력 스펙트럼 밀도는 다음 식으로 구할 수 있다.

$$U_0 = \mathbf{A}^H \mathbf{x} = \mathbf{A}^H \mathbf{x}^s + \mathbf{A}^H \mathbf{x}^n \quad (30)$$

$$\mathbf{U}_{1:M-1} = \mathbf{B}^H \mathbf{x} = \mathbf{B}^H \mathbf{x}^s + \mathbf{B}^H \mathbf{x}^n \quad (31)$$

$$P_{U_0} = \mathbf{A}^H \mathbf{P}_X^s \mathbf{I}^s \mathbf{A} + \mathbf{A}^H \mathbf{P}_X^n \mathbf{I}^n \mathbf{A} = P_X^s + \frac{P_X^n}{M} \quad (32)$$

$$\mathbf{P}_{\mathbf{U}_{1:M-1}} = \mathbf{B}^H \mathbf{P}_X^s \mathbf{I}^s \mathbf{B} + \mathbf{B}^H \mathbf{P}_X^n \mathbf{I}^n \mathbf{B} = P_X^n \mathbf{I}. \quad (33)$$

이 때, 차단 행렬은 유니타리 행렬을 선택한다. 위 식들로부터 a priori SNR의 추정값은 다음과 같이 구할 수 있다.

$$\gamma \approx \frac{P_{U_0}}{(\text{Tr}\{\mathbf{P}_{\mathbf{U}_{1:M-1}}\})/(M-1)} - \frac{1}{M}. \quad (34)$$

IV. 실험 결과

제안하는 알고리즘을 검증하기 위한 실험을 위해서 image method를 이용한 가상의 데이터를 사용하였다^[4]. 방의 크기를 가로 6m, 세로 4m, 높이 2.5m로 정한 후, 그림2와 같이 높이 1m 위치에 4개의 마이크를 20cm 등간격으로 배치한 후 정면에서 1.5m 떨어진 곳에서 높이 1m에서 음성이 발생되며 잡음은 백색잡음으로서 정면과 63°의 각도를 이루며 높이 1m, 거리 2.2m 떨어진 곳에서 마이크에 도달하게 된다. 3~4초 정도 길이의 음성데이터 20개를 입력으로 넣고 SNR이 0 dB가 되도록 잡음의 크기를 조정하였다. 본 실험에서는 여러 음성 신호처리에서 특징벡터로 널리 사용되는 MFCC (Mel-Frequency Cepstral Coefficient)의 거리를 평가척도로 하였다. MFCC의 차수는 음성인식에서 혼

히 쓰이는 13차로 하였다. 제안하는 알고리즘을 적용하여 잡음을 제거한 신호와 가장 왼쪽 마이크로폰에 도달하는 음성 신호 (잡음이 섞이지 않은)와의 MFCC 거리를 측정하였다. 표 1에 잡음 제거 필터를 통과하지 않은 신호의 MFCC 거리, 가중치를 1로 설정한 잡음 제거 필터를 통과한 신호와의 MFCC 거리, 제안하는 알고리즘에 따라 가중치를 설정한 잡음 제거 필터를 통과한 신호와의 MFCC 거리를 나타내었다. 실험 결과 제안하는 방법에 따라 SP-SDW-MWF의 가중치를 적용함으로써 목적 신호와 추정 신호간의 MFCC 거리를 줄일 수 있음을 알 수 있다.

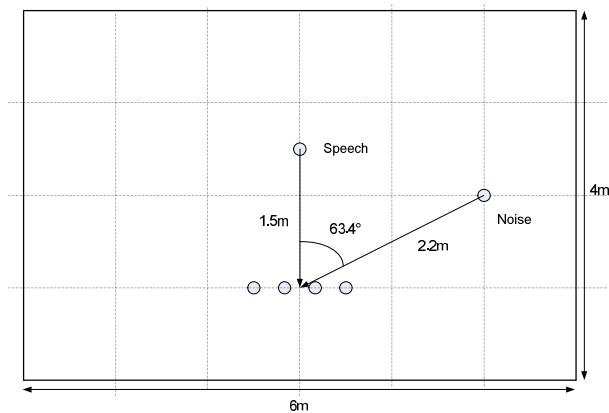


그림 2. 실험환경
 Fig. 2. Simulation environment

표 1. MFCC 거리
 Table 1. MFCC distance

	잡음 제거 전	가중치 1을 이용한 잡음 제거 필터 통과	제안하는 방법을 이용하여 가중치를 설정한 잡음 제거 필터 통과
MFCC 거리	4.29	0.753	0.686

V. 결론

본 논문에서는 SP-SDW-MWF를 이용한 어레이 마이크로폰 잡음 제거 시스템에서 음성 왜곡 가중치를 결정하는 방법을 제안하였다. 시스템의 출력 신호와 목적 신호의 전력 스펙트럼 밀도 오차로 성능을 평가하는 경우, 가중치는

a priori SNR에 따라 다른 값을 가져야 함을 보였다. 고정 빔포머와 차단 행렬의 출력을 이용하여 a priori SNR을 추정하고, 그에 따른 가중치를 적용하였다. MFCC 거리를 측정하는 실험에서 a priori SNR 추정값에 따른 가중치를 사용하여 보다 나은 성능을 얻을 수 있음을 알 수 있다. 본 논문에서 제안하는 알고리즘은 입력 신호의 a priori SNR 추정의 정확도에 따라 그 성능이 좌우될 수밖에 없는데 a priori SNR의 추정은 여러 음성 신호 처리에서 중요하게 다루어지는 분야이면서 잡음 환경에 따라 오차의 정도가 다양하게 나타나고 있다. 본 실험에서는 비교적 a priori SNR의 추정 오차가 적은 백색잡음에 대해서만 실험결과를 표시하였다. 향후 다채널 입력 상황에서 다양한 a priori SNR 추정 방법을 이용하여 여러 잡음 환경에서 실험을 추가하여 보다 실용적인 알고리즘으로 발전시키고자 한다.

참고 문헌

- [1] J. S. Lim and a. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proceedings of the IEEE, vol. 67, no. 12, pp. 1586 - 1604, 1979.
- [2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-27, no. 2, pp. 113 - 120, 1979.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-32, no. 6, pp. 1109 - 1121, 1984.
- [4] Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. on Speech and Audio Processing, vol. 3, no. 4, pp. 251 - 266, 1995.
- [5] J. Huang and Y. Zhao, "An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises," Speech Communication, vol. 26, no. 3, pp. 165 - 181, Nov. 1998.
- [6] K. Hermus, P. Wambacq, and H. Hamme, "A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition," EURASIP Journal on Advances in Signal Processing, vol. 2007, no. 1, p. 045821, 2007.
- [7] M. Brandstein and D. Ward (Eds.), Microphone Arrays, Springer-Verlag, 2001.
- [8] S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," IEEE Trans. on Signal Processing, vol. 50, no. 9, pp. 2230-2244, Sep. 2002.
- [9] G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," Signal Processing, vol. 83, no. 9,

pp. 1889-1904, Sep. 2003.

- [10] A. Spriet, M. Moonen, and J. Wouters, "Spatially preprocessed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367-2387, Dec. 2004.
- [11] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagation*, vol. AP-30, no. 1, pp. 27-34, Jan. 1982.

- [12] J. Benesty, S. Makino, J. Chen (Eds.), *Speech Enhancement*, Springer, 2005.
- [13] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.
- [14] J. Allen and D. Berkely, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943-950, April 1979.

저 자 소 개



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
- 2010년 ~ 2011년 : 대구대학교 전자공학부 전임강사
- 2011년 ~ 현재 : 송실대학교 조교수
- 주관심분야 : 음성신호처리, 영상신호처리, 멀티모달신호처리, 어레이신호처리