

# 마이크로폰 어레이 시스템을 사용한 블라인드 처리에 의한 음원분리

박진남\*, Usagawa Tsuyoshi\*\*, Masanao Ebata\*\*

## Source signal separation by blind processing for a microphone array system

Jin-Nam Park\*, Usagawa Tsuyoshi\*\*, Masanao Ebata\*\*

\* : 한국전기연구소(KERI), \*\* : Kumamoto University(Japan)

E-mail : jnpark@keri.re.kr

### 요약

본 논문에서는 음원에 관한 정보가 미지의 상황에서 마이크로폰 어레이를 사용하여 두 음원신호를 분리하는 시스템을 제안한다. 이 시스템은 두 단계로 구성되어 있으며, 첫 번째 단계에서는 파워가 큰 제 1음원의 DOA(Direction Of Arrival)를 추정하고, AMUSE (Algorithm for Multiple Unknown Signals Extraction) 법을 사용한 Blind Deconvolution에 의해 음원신호의 분리를 행한다. 두 번째 단계에서는 파워가 낮은 제 2음원의 강조신호를 사용하여 DSA(Delay and Sum Array)법에 의해 제 2음원의 DOA를 추정하고, AMUSE법의 출력신호와 두 음원의 DOA를 이용하여 ANF(Adaptive Notch Filter)를 구성하고, 두 음원신호의 재 분리를 행한다. 그리고, 시뮬레이션을 통해 제안한 방법의 유효성을 검토한 결과 두 음원 신호가 분리 가능한 것이 확인되었다.

### 1. 서론

최근 급속한 멀티미디어의 발전과 함께 통신시스템을 이용한 원격 통신회의 시스템과 같은 서로 다른 곳에서 마이크로폰과 영상을 이용한 통신 시스템의 수요가 늘어나고 있다. 이 같은 시스템의 수요에 따라 질 좋은 원격 통신시스템에 관한 요구도 급증하고 있다. 그러나, 마이크로폰을 사용하여 음성을 수신하는 경우, 소음 및 방해 음성 등의 잡음에 의해 음질의 열화가 현저하게 나타나므로 이들 잡음 성분을 되도록 수신하지 않도록 하여 높은 SNR(Signal to Noise Ratio)을 갖는 수신기술 개발이 필요하다. 이 기술로써는 마이크로폰 어레이를 사용한 신호분리 시스템이 유효한 방법 중에 하나로 인식되어 활발한 연구가 진행중이다.

마이크로폰 어레이를 사용한 신호분리 시스템에 관한 연구는 다수 제안되어져 왔지만, 이들 방법의 대부분은 음원간의 신호세기의 차이에 관한 제한 및 음원의 수와 마이크로폰의 수에 관한 제한 등 여러 가지 제약점을 가진 방법들이 많이 제안되어져 왔다. 그러나 최근 음원의 수와 마이크로폰의 수에 크게 의존하지 않으면서 음원에 대한 정보가 없이 원하는 목적 음원을 추정하는 Blind Deconvolution방법이 심도 깊게 연구되어지고 있다.

본 논문에서는 미지의 음원이 존재하는 환경 하에서 마이크로폰 어레이를 사용한 미지의 두 음원 신호 분리 시스템을 제안하고 그 유효성을 검토한 결과에 대해 기술한다. 본 시스템의 유효성을 검토한 결과 미지의 두 음원이 분리 가능한 것을 확인할 수 있었다.

### 2. Blind Deconvolution을 이용한 두 음원 분리 시스템

음원의 방향 및 음원 특성에 관한 정보가 미지의 상황에서 두 음원 분리를 시험하는 시스템을 제안한다. 그림 1은 이 방법의 block diagram을 나타낸다. 그림 1과 같이 이 방법은 크게 두 단계로 나누어져 있다. 첫 번째 단계는 먼저 DSA법에 의해 최대 파워를 가진 음원의 DOA를 추정한다. 또, 수신신호를 복수의 sub-array로 분해하여 sub-array단위로 AMUSE법을 사용한 Blind deconvolution을 행하여 각 음원신호의 강조를 행한다. 여기서 각 AMUSE에 의한 출력신호를 각 음원의 강조신호  $y_1$ 과  $y_2$ 를 그룹화 시켜 두 번째 단계의 입력신호로 사용한다. 두 번째 단계에서는 AMUSE법의 출력신호를 사용하여 DSA법에 의한 제 2음원의 DOA를 추정하고, AMUSE법의 출력신호와 각 음원신호의 DOA를 바탕으로 구성된 ANF를 사용하여

음원을 분리한다. 이 절에서는 제안하는 방법의 알고리즘에 관해서 기술한다.

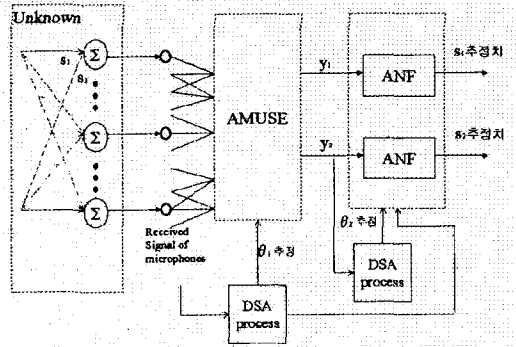


그림 1. 두 음원 분리시스템 블록 다이어그램

## 2.1 마이크로폰 어레이 출력신호의 표현

그림 2와 같이,  $m_1 \sim m_M$ 의  $M$ 개의 마이크로폰을 등간격  $d$ 로 직선 상에 배치하고, 방향이 다른 2개의 음원을 관측하는 직선 마이크로폰 어레이 시스템이고, 각 음원신호는 평면파로 마이크로폰에 도달하며, 반사는 생기지 않는다고 가정한다. 마이크로폰  $m_i$ 에서 수신된 2개의 음원 중에 파워가 큰 신호(제 1음원)를  $s_1(k\Delta T)$ , 다른 하나의 신호(제 2음원)를  $s_2(k\Delta T)$ 로 하고 음원의 방향은 각각  $\theta_1, \theta_2$ 로 하고,  $\Delta T$ 는 샘플링 간격을 나타내고, 마이크로폰  $m_i$ 의 부가잡음을  $n_i(k\Delta T)$ 로 할 때 마이크로폰  $m_i$ 의 출력신호  $x_i(k\Delta T)$

$$x_i(k\Delta T) = s_1(k\Delta T - \tau_{\theta_1, i}) + s_2(k\Delta T - \tau_{\theta_2, i}) + n_i(k\Delta T) \quad (1)$$

이다. 단,  $\tau_{\theta, i}$ 는 마이크로폰  $m_i$ 을 기준으로 방향각  $\theta$ 에 대한 마이크로폰  $m_i$ 의 시간 지연이다.

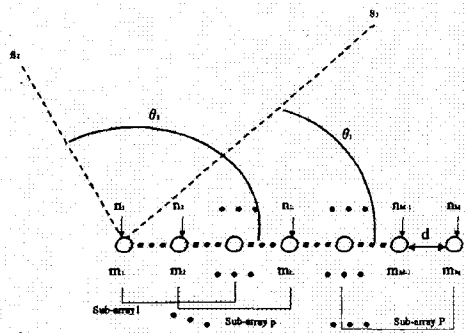


그림 2 직선 마이크로폰 어레이 배치도

## 2.2 Blind Deconvolution에 의한 두음원신호 분리

그림 1의 AMUSE 단계에서는 DSA법에 의한 최대 파워를 가진 음원의 방향 추정 및 AMUSE법을 사용한

Blind Deconvolution에 의한 음원신호의 분리를 행한다. 마이크로폰 어레이의 수신신호를 사용하여 최대 파워를 가진 음원의 방향을 추정한다. DSA법의 파워 추정치  $P_{DSA}(\theta)$ 는

$$\begin{aligned} P_{DSA}(\theta) &= \frac{1}{K} \sum_{k=1}^K \left| \sum_{i=1}^M e^{-j2\pi R(i-1)d \sin \theta / C} \cdot x_i(k\Delta T) \right|^2 \\ &= \frac{1}{K} \sum_{k=1}^K |d(\theta)^H x(k\Delta T)|^2 \quad (2) \\ &= d(\theta)^H \widehat{R}_x d(\theta) \quad (-\pi/2 \leq \theta \leq \pi/2) \end{aligned}$$

이다.  $x(k\Delta T)$ 는 마이크로폰 어레이 수신신호벡터로  $x(k\Delta T) = [x_1(k\Delta T), \dots, x_M(k\Delta T)]$ 이다.  $C$ 는 음속을 나타내고,  $H$ 는 복소수 공액전치,  $d(\theta)$ 는 방향 탐색 복소수 벡터,  $\widehat{R}_x$ 는  $x(k\Delta T)$ 의 공분산 행렬이다. 식 (2)에 정의된  $P_{DSA}(\theta)$ 의 값이 최대가 되었을 때의  $\widehat{\theta}_1$ 를 제 1음원의 DOA라 한다. 마이크로폰 수신신호  $x_i(k\Delta T)$ 를  $P$ 개의 sub-array로 분해하고, sub-array 단위로 AMUSE법에 적용시켜 Blind Deconvolution을 행한다. 각각의 sub-array의 출력신호는 벡터  $x_p(k\Delta T)$  ( $p=1, \dots, P$ )로 표현한다.

$$x_p(k\Delta T) = [x_p(k\Delta T), \dots, x_{p+L-1}(k\Delta T)] \quad (3)$$

$L$ 은 sub-array를 구성하는 마이크로폰의 수이고, 추정된  $\widehat{\theta}_1$ 방향신호를 sub-array 단위로 동기화 시키기 위해 시간 지연  $\tau_{\widehat{\theta}_1, i}$ 를 추가한다.

$$\widetilde{x}_p(k\Delta T) = [x_p(k\Delta T), x_{p+1}(k\Delta T + \tau_{\widehat{\theta}_1, p+1} - \tau_{\widehat{\theta}_1, p}), \dots, x_{p+L-1}(k\Delta T + \tau_{\widehat{\theta}_1, p+L-1} - \tau_{\widehat{\theta}_1, p})] \quad (4)$$

$\widetilde{x}_p(k\Delta T)$ 의 공분산 행렬  $\widehat{R}_{\widetilde{x}_p}$ 의 고유벡터  $u_{p, l}$ 와 고유치  $\lambda_{p, l}$  ( $l=1, \dots, L$ )는

$$\widehat{R}_{\widetilde{x}_p} = E[\widetilde{x}_p(k\Delta T) \widetilde{x}_p(k\Delta T)^T]$$

$$SVD(\widehat{R}_{\widetilde{x}_p}) = [u_{p, 1}, \dots, u_{p, L}] \text{diag}(\lambda_{p, 1}^2, \dots, \lambda_{p, L}^2) [u_{p, 1}, \dots, u_{p, L}]^T \quad (5)$$

이다.  $SVD(\cdot)$ 는 고유치 분해,  $T$ 는 전치를 나타내며, 고유치  $\lambda_{p, l}$ 가 음원의 파워에 의존하는 특성을 이용하여 어떤 임계치를 넘는  $\lambda_{p, l}$ 의 값을 음원수의 추정치이고, 음원에 대한 특성 계수  $q_p(k\Delta T)$ 는 다음과 같다.

$$q_p(k\Delta T) = T_p \widetilde{x}_p(k\Delta T) \quad (6)$$

$$T_p = \text{diag}\left(\frac{1}{d_1}, \frac{1}{d_2}, \dots, \frac{1}{d_N}\right) U_{p, s}^T$$

$$d_l = \sqrt{\lambda_{p, l}^2 - \sigma^2}, \quad l=1, \dots, \widehat{N} \quad (7)$$

$$U_{p, s}^T = [u_{p, 1}, \dots, u_{p, \widehat{N}}]$$

$\sigma^2$ 는 부가잡음에 대응하는 분산이고,  $\widehat{N}$ 은 음원수의 추정치이다.  $q_p(k\Delta T)$ 의 공분산 행렬  $\widehat{R}_{q_p}$ 의 SVD에 의한 고유벡터  $V_p(k)$ 는

$$V_p \Sigma_p V_p^T = SVD((\widehat{R}_{q_p} + \widehat{R}_{q_p}^T)/2) \quad (8)$$

이다.  $\Sigma_p$ 는 특이치행렬이고, 고유벡터  $V_p$ 를 사용하여 각 sub-array의 음원의 추정치  $z_p(k\Delta T)$ 를 계산한다.

$$\mathbf{z}_p(k\Delta T) = \mathbf{V}_p^T \mathbf{q}_p(k\Delta T) \quad (9)$$

추정된 음원의 수가 1개 이하일 때  $\mathbf{z}_p(k\Delta T)$ 를 본 방법의 출력신호로 하지만, 음원의 수를 2개 이상으로 추정된 경우는 sub-array의 출력 신호를 각 음원의 강조신호별로 재배열하고, 두 번째 단계로 출력한다. 다음 식에서 표현되는  $\hat{\theta}_1$ 방향의 강조신호의 벡터  $\hat{\mathbf{y}}_j(k\Delta T)$ 를 두 번째 단계의 출력신호로 한다.

$$\hat{\mathbf{y}}_j(k\Delta T) = [y_{j,1}(k\Delta T), \dots, y_{j,p}(k\Delta T), \dots, y_{j,p}(k\Delta T), \dots] \quad (10)$$

$y_{j,p}(k\Delta T)$ 는  $p$ 번째의 sub-array에 있어서  $\theta_j$ 방향의 신호성분을 강조한 신호이다.

### 2.3 Adaptive notch filter에 의한 두 음원신호의 재 분리

두 번째 단계 ANF에서는 첫 번째 단계의 출력신호  $\mathbf{y}_2(k\Delta T)$ 에 DSA법을 적용시켜 제 2음원의 DOA  $\hat{\theta}_2$ 를 추정하고 AMUSE의 출력신호와 DSA법에 의해 추정된 각 음원의 DOA를 기준으로 음원분리를 행한다.

AMUSE의 출력신호  $\mathbf{y}_j(k\Delta T)$ 은 입력신호  $s_j(k\Delta T)$ 를 강조한 신호이다.

$$y_{j,p}(k\Delta T) = \hat{s}_1(k\Delta T - \tau_{\hat{\theta}_1,p}) + \hat{s}_2(k\Delta T - \tau_{\hat{\theta}_2,p}) \quad (11)$$

$\hat{s}_1(k\Delta T)$ 를 추출하는 경우  $\hat{s}_2(k\Delta T)$ 는 잡음으로 간주되고,  $\hat{s}_2(k\Delta T)$ 를 추출하는 경우에는  $\hat{s}_1(k\Delta T)$ 가 잡음으로 간주된다. 우선,  $\hat{s}_1(k\Delta T)$ 를 바라는 음원신호로,  $\hat{s}_2(k\Delta T)$ 를 잡음으로 하는 경우에 대해서 살펴보면,  $\hat{\theta}_1$ 방향으로 감도를 높이기 위해서 동기가산을 한다.

$$s_{\text{sum}}(k\Delta T) = \sum_{p=1}^P y_{1,p}(k\Delta T + \tau_{\hat{\theta}_1,p}) \quad (12)$$

$$= P\hat{s}_1(k\Delta T) + \sum_{p=1}^P \hat{s}_2(k\Delta T - \tau_{\hat{\theta}_2,p} + \tau_{\hat{\theta}_1,p})$$

식(12)의 양변을  $z$ 변환하면 다음 식과 같이 된다.

$$S_{\text{sum}}(z) = P\hat{S}_1(z) + g(z)\hat{S}_2(z) \quad (13)$$

$$g(z) = \sum_{p=1}^P z^{(-\tau_{\hat{\theta}_2,p} + \tau_{\hat{\theta}_1,p})/f_s}$$

단,  $f_s$ 는 샘플 주파수이다. sub-array 1과 sub-array  $p$  ( $p \geq 2$ )의 출력신호로부터 얻으며,  $\hat{s}_1(k\Delta T)$ 의 성분을 포함하고 있지 않는 신호성분  $\tilde{s}_{1,p}(k\Delta T)$ 는

$$\tilde{s}_{1,p}(k\Delta T) = y_{1,1}(k\Delta T) - y_{1,p}(k\Delta T + \tau_{\hat{\theta}_1,p}) \quad (14)$$

$$= \hat{s}_2(k\Delta T) - \hat{s}_2(k\Delta T - \tau_{\hat{\theta}_2,p} + \tau_{\hat{\theta}_1,p})$$

이고, 양변을  $z$ 변환하면 다음 식과 같이 된다.

$$\tilde{S}_{1,p}(z) = h_p(z)\hat{S}_2(z), \quad h_p(z) = 1 - z^{(-\tau_{\hat{\theta}_2,p} + \tau_{\hat{\theta}_1,p})/f_s} \quad (15)$$

또한,  $\tilde{s}_1(z)$ ,  $h_p(z)$ 를 요소로 하는 벡터를  $\mathbf{h}(z)$ 라고 표현하면 다음 식과 같이 수정할 수 있다.

$$\tilde{\mathbf{s}}_1(z) = \mathbf{h}(z)\hat{\mathbf{S}}_2(z) \quad (16)$$

$$\tilde{\mathbf{s}}_1(z) \equiv [\tilde{S}_{1,1}(z), \dots, \tilde{S}_{1,p}(z), \dots, \tilde{S}_{1,P}(z)]^T$$

$$\tilde{\mathbf{h}}(z) \equiv [h_1(z), \dots, h_p(z), \dots, h_P(z)]^T$$

식(13)과 식(16)으로부터  $\hat{\theta}_1$ 방향의 신호추정치  $\hat{S}_1(z)$ 는  $\hat{S}_1(z) = \frac{1}{P}(S_{\text{sum}}(z) - g(z)\mathbf{h}^{-1}(z)\tilde{\mathbf{s}}_1(z))$  (17)로 주어진다. 따라서, 형식적으로는 식(17)의 양변을 역  $z$ 변환하는 것에 의해  $\hat{s}_1(k)$ 를 얻는 것이 가능하다. 또한 음원신호  $s_2(k)$ 의 추정치  $\hat{s}_2(k)$ 도  $\hat{s}_1(k)$ 의 추정방법과 같은 방법으로 추정할 수 있다. 그러나, 식(17)은  $\mathbf{h}(z)$ 의 역 필터를 구할 필요가 있지만,  $\mathbf{h}(z)$ 는 단순한 지연을 포함하는 형이기 때문에 복수의 영점을 가지고, 직접적인 방법에서는 안정한 역 필터를 구하는 것이 힘들다. 이 문제를 해결하기 위해 역 필터를 구하기 위해 반복형 에코 제거법[3]을 적용하여 계산한다.

### 3. 검증 시뮬레이션

제한한 방법의 유효성 테스트를 위하여 계산기 시뮬레이션을 하였다. 시뮬레이션에서는 6개의 마이크로폰을 직선 상에 등간격  $d=0.08m$ 로 배열한 직선 마이크로폰 어레이를 사용하였다. 이때, sub-array 수는 3개, sub-array의 마이크로폰 수는 4개로 한 경우에 대하여 검토한다. 이 이후에 기술하는 방향각은 마이크로폰  $m_1$ 을 중심으로 마이크로폰 어레이의 축방향으로부터 반시계방향으로의 각도를 표시한다. 또, 샘플주파수는  $10kHz$ , 반복형 에코제거 레벨은  $L_{th}=40$ 으로 한다. [3] 수신신호는 평면파로 가정하고, 반사는 생기지 않는 것으로 한다.

제한한 방법의 분리 성능을 테스트하기 위해 음성  $/a/$ 에 대하여 음성  $/u/$ 의 파워 비를  $-20dB$ 부터  $20dB$ 까지 변화를 시킨 경우 각 음원의 추정신호를  $300Hz$ 부터  $3000Hz$ 까지의 주파수 대역에 있어서의 SNR을 계산하였다. 이 결과 파워가 큰 쪽의 신호(제1음원)의 추정 성능이 좋지만, 이것에 비하여 제2음원의 추정 성능은 그리 좋지 못하다. 또, 제한한 방법이 부가잡음에 대해서는 영향을 받지 않는 특성이 있음을 논문[4]를 통해 기술했다. 또한, 그림 3에는 음원1로써 일본인 화자가 발음한 음성신호 “저쪽에 사거리 신호가 있습니다”를  $120^\circ$ , 음원2는 음원1에 대하여  $-10dB$ 의 파워를 가진 음성신호 “회장은 어느 쪽입니까?”를  $30^\circ$ 에 배치하여 시뮬레이션을 행한 결과 나타낸다. 그림 3의 파형은 위에서부터 각 음성신호의 원음, 마이크로폰 1의 출력신호 및 추정신호를 시간특성으로 나타낸 결과이다. 음원2의 추정신호에서 뒤쪽 끝부분이 추정되지 않았지만, 이것은 음원 1의 파워에 대해 제2음원의 파워가 아주 작기 때문에 AMUSE의 단계에서 음원 수를 1로 잘못 추정하였기 때문이다. 그림 4에는 그림3의 음성신호의 주파수 특성을 보이고 있다. 주파수 영역에서도 두

신호가 유사도 높게 분리되어 있음을 확인할 수 있다.

#### 4. 맺음말

AMUSE에 의한 Blind Deconvolution을 사용한 음원강조와 Adaptive notch filter를 응용한 음원분리 방식을 조합하여 두음원 분리 시스템을 제안하였다. 시뮬레이션 결과 두 음원분리가 가능하였고, 부가잡음에 대해서도 영향을 받지 않는 시스템임을 알 수 있었다. 그러나, 향후 연구 과제로써 음원추정 성능에 대한 평가 지수로써 사람이 인지하는 유사도 측면을 강조하는 평가함수의 개발과 파워가 작은 음원의 추정성능을 향상시킬 필요성이 있다.

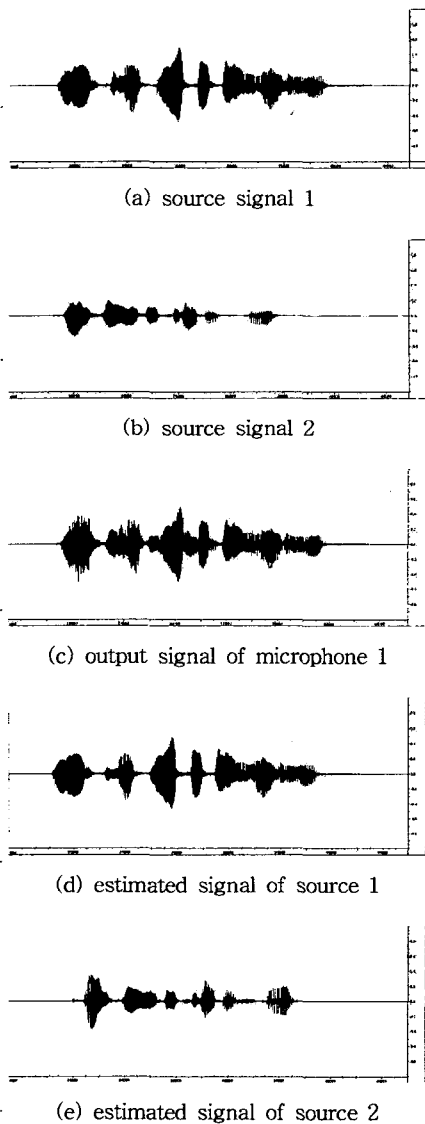


그림 3 음성신호의 추정결과 (시간특성)

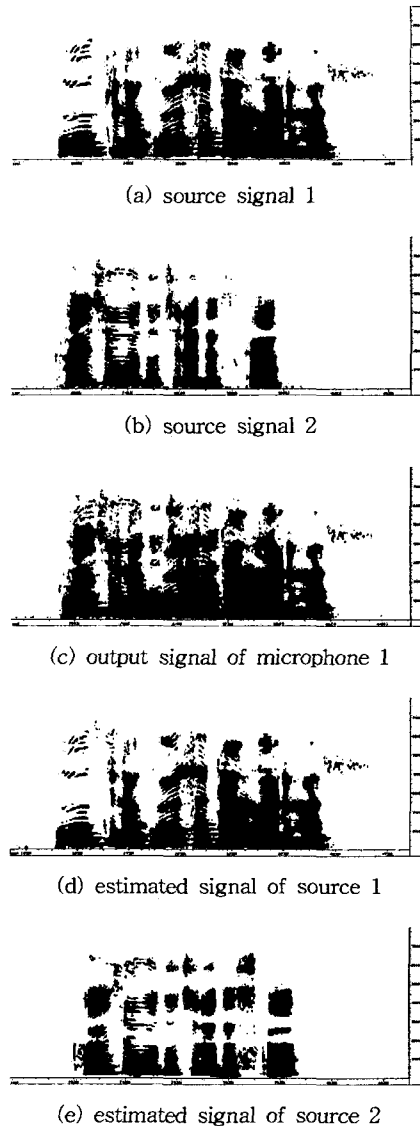


그림 4 음성신호의 추정결과 (주파수특성)

#### 참고 문헌

- [1] Simon Haykin, "Blind Deconvolution," Prentice-Hall, Englewood Cliffs, NJ, 1994
- [2] Lang Tong, Ruy-wen Liu, "Indeterminacy and Identifiability of Blind Identification," vol.38, no.5, pp.499-509, IEEE Trans. on CAS, 1991
- [3] H. Kouda, "계의 영점을 고려한 반복형 에코 제거법," 일본음향학회지, 53권 4호, pp.268-276, 1997
- [4] J.N. Park, T. Usagawa and M. Ebata, "A speech enhancement method using an iterative system based on blind deconvolution," pp. 45-50 (EA97-67), Technical Report of IEICE, Nov.,1998