

## DOA 기반 학습률 조절을 이용한 다채널 음성개선 알고리즘

### Multi-Channel Speech Enhancement Algorithm Using DOA-based Learning Rate Control

김 수 환<sup>1)</sup> · 이 영 재 · 김 영 일 · 정 상 배<sup>2)</sup>

Kim, Suhwan · Lee, Youngjae · Kim, Young-Il · Jeong, Sangbae

#### ABSTRACT

In this paper, a multi-channel speech enhancement method using the linearly constrained minimum variance (LCMV) algorithm and a variable learning rate control is proposed. To control the learning rate for adaptive filters of the LCMV algorithm, the direction of arrival (DOA) is measured for each short-time input signal and the likelihood function of the target speech presence is estimated to control the filter learning rate. Using the likelihood measure, the learning rate is increased during the pure noise interval and decreased during the target speech interval. To optimize the parameter of the mapping function between the likelihood value and the corresponding learning rate, an exhaustive search is performed using the Bark's scale distortion (BSD) as the performance index. Experimental results show that the proposed algorithm outperforms the conventional LCMV with fixed learning rate in the BSD by around 1.5 dB.

**Keywords:** Speech enhancement, linear constrained minimum variance, direction of arrival, adaptive learning rate control, beamforming

#### 1. 서 론

기계-인간 간의 인터페이스를 위하여, 음성은 사용자들에게 가장 편안하고 자연스러운 수단이기 때문에 음성 인터페이스들은 널리 연구되어 왔다. 그러나 음성 인터페이스의 성능은 주변 환경 잡음에 의해서 심각하게 저하되는 경향이 있다. 이러한 이유 때문에, 음성 인터페이스의 성능 개선을 위하여 다양한 음성 개선 기술들이 제안되어 왔다. 잡음의 종류에는 정상성 잡음과 비정상성 잡음으로 나눌 수 있다. 잡음의 특성에 따라서 사용해야 하는 잡음 제거 기술도 달라진다. 일반적으로 정상성 잡음을 제거하는 방법으로는 단일채널을 기반으로 하는 Wiener 필터와 Kalman 필터 방법이 있다[1][2]. 반면에 비정상성 잡음을 제거하기 위한 방법으로는 다채널을 기반으로 하는 빔포밍과 blind signal separation(BSS)가 있다[3-6][7][8]. 보통의 빔포

밍 기술은 주로 TV나 로봇을 위한 음성 인터페이스에서 적용을 위한 연구가 수행되어 왔으며 4-채널 이상을 사용하고 있다. 반면에 이동형 단말기를 위한 빔포밍 기법은 하드웨어 상의 제약 때문에 2-채널 이상을 사용하기 어렵다. 빔포밍과 BSS의 차이점은 다음과 같다. BSS는 실시간 계산량의 문제와 사용하는 마이크로폰의 개수가 분리해야 하는 음원의 개수에 의존하는 문제 때문에 실제로 적용할 때 좋은 성능을 발휘하는데 어려움이 따른다[4][6]. 반면에, 빔포밍은 목표 음성신호의 위치정보를 미리 알 수 있기 때문에 BSS보다 실제 환경에서 더 좋은 성능을 발휘함이 알려져 있다[10]. 가장 널리 사용되는 빔포밍 알고리즘으로 LCMV와 generalized sidelobe canceller (GSC)가 있다[3][5][7-9]. GSC는 목표 음성신호 구간에서는 잡음 제거를 수행하지 않기 때문에 마이크로폰의 개수를 적게 사용할 수밖에 없는 이동형 단말기에서는 좋은 성능을 발휘하기 어렵다. 하지만 LCMV의 경우에는 목표 음성신호 구간에서도 잡음 제거를 수행하기 때문에 마이크로폰의 개수가 적을 때 음질의 측면에서는 GSC 보다 더 나은 성능을 보인다고 볼 수 있다.

본 논문에서 제안하고 있는 음성 개선 알고리즘은 휴대단말기에 적용하는 것이 목적이다. 그러므로 2-채널의 마이크로폰을 사용하였고 목표 음성 신호 구간에서 음질을 개선하는 것이 목

1) 경상대학교 전자공학과, edps2166@gnu.ac.kr

2) 경상대학교 전자공학과(공학연구원), jeongsb@gnu.ac.kr, 교신저자

접수일자: 2011년 8월 3일

수정일자: 2011년 9월 21일

게재결정: 2011년 9월 23일

표이기 때문에 참고문헌 [11]의 LCMV를 기반으로 하였다. 일반적인 LCMV는 목표 음성신호 구간이나 잡음 구간에 상관없이 항상 같은 학습률로 잡음 제거를 수행한다. 목표 음성신호-잡음 구간을 구별하지 않고 같은 학습률로 잡음 제거를 하였을 경우, 학습률이 높을 때에는 배경잡음을 많이 제거할 수 있으나 상대적으로 높은 signal-to-noise ratio (SNR)을 갖는 목표 음성의 일부 구간에서는 과도한 잡음제거로 신호의 왜곡을 초래할 수 있다. 만약, 학습률이 낮으면, 목표 음성 구간의 왜곡을 방지할 수 있는 반면 순수 배경 잡음 구간에서의 잡음제거의 양이 적을 수 있다[9]. 이러한 문제점을 해결하기 위해서 본 논문에서는 DOA 기반의 적응적 학습률 변환 기법을 기존의 LCMV에 적용하였다. 이를 위해서, 먼저 2-채널 기반의 단구간 신호로부터 DOA를 추정하고 그것을 이용하여 목표 음성의 존재 우도를 측정하였다. 최종적으로 측정된 우도를 바탕으로 LCMV 내의 적응 필터의 학습률을 결정하여 적용하였다.

본 논문의 구성은 다음과 같다. 제 2장에서 관련 연구를 소개하고, 제 3장에서 제안된 알고리즘에 대해서 설명한다. 제 4장에서 실험 결과에 대해서 논하고 마지막으로 제 5장에서 본 논문의 결론을 맺는다.

## 2. 관련 연구

빔포밍은 인간의 음성 및 음악신호와 같은 비정상성 잡음을 방향정보를 이용하여 제거한다. 빔포밍 기술은 크게 LCMV와 GSC 알고리즘이 연구되어 왔으며 각각 서로 다른 관점에 의해서 방향성 잡음을 제거한다. 각 알고리즘에 대한 자세한 설명은 다음과 같다.

### 2.1 Generalized Sidelobe Canceller

GSC는 전형적인 적응적 빔포밍이다. 그리고 실제로 적용이 많이 되고 있는 빔포밍 기법 중에 하나이다. GSC는 고정 빔포머부, 차단행렬부, 잡음 제거부로 구성된다. 고정 빔포머부에서는 목표 음성 신호원에 대해서 각 채널의 이득 및 시간 지연이 같아지도록 하기 때에는 목표 음성신호 구간에서는 전력(power)이 최대가 되고 잡음 구간에서는 서로 상쇄되어 상대적으로 전력을 줄여주는 역할을 수행한다. 차단행렬부에서는 목표 음성 신호원의 방향 정보를 이용하여 목표 음성 신호를 차단하여 참조 잡음신호만을 추출하는 역할을 수행한다. 만일, 차단행렬부에서 목표 신호를 제대로 차단하지 못하면 목표 음성 신호 부분까지 잡음 제거가 수행되기 때문에 빔포밍 결과에서 목표 음성신호가 왜곡된 결과를 얻을 수 있다. 잡음 제거부에서는 차단행렬부에서 출력된 잡음을 고정 빔포밍 결과에 잔존하고 있는 잡음과 유사하도록 적응적으로 필터링하여 잡음을 제거하는 역할을 수행하게 된다.

### 2.2 Linearly Constrained Minimum Variance

LCMV 또한 전형적인 적응적 빔포밍이다. LCMV의 기본 원리는 정면으로 선형조건을 두어 출력의 전력을 최소화 하는 것이다. 그림 1은 LCMV 빔포머의 기본적인 구조를 나타내고 있다.

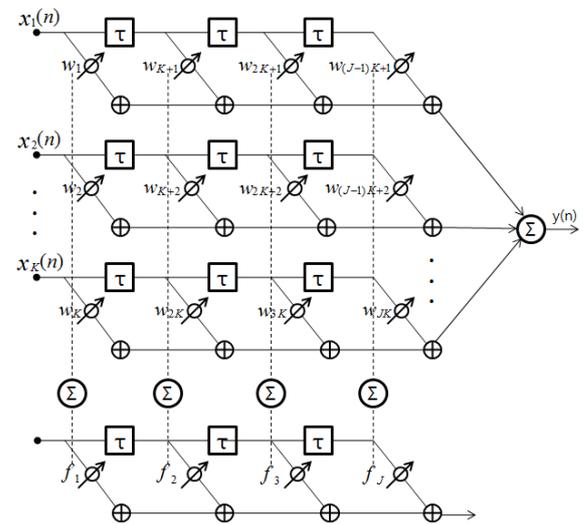


그림 1. LCMV 빔포머의 기본적인 구조  
Figure 1. The framework of LCMV beamformer

그림 1에서 K와 J는 각각 마이크로폰의 개수와 적응적 필터의 길이이다. LCMV 알고리즘에서는 목표 음원은 마이크로폰 배열의 정면에 위치된다고 가정한다. 목표 음성신호의 방향이 아닌 다른 방향의 신호는 잡음으로 간주한다. 그림 1의 맨 아래 부분은 목표 음성신호원에 대한 LCMV 빔포머의 등가 구조를 보여주고 있다. LCMV 빔포밍의 기본 방식은 빔포머의 출력을 최소화하는 동시에 목표 음성 신호의 주파수 응답은 항상 1이 되도록 하는 것이다. 이러한 동작을 수식 (1)에 나타내었다.

$$\min_w E[y^2(n)] \quad f(n) = \delta(n) \quad (1)$$

여기서  $\vec{w} = [w_1, w_2, \dots, w_K]$ 이고,  $E[\cdot]$ 는 통계적 평균 연산자이다.  $y(n)$ 과  $f(n)$ 은 각각 빔포머의 출력과 등가 필터가 중치이다. 수식 (1)은 Lagrange multiplier에 의해 최적화 된다. 최종적으로 수식 (2)로 정리된다[11].

$$W_{opt} = R_{xx}^{-1} C [C^T R_{xx} C]^{-1} F \quad (2)$$

여기서  $R_{xx}$ 는  $\vec{x} = [x_1^T(n), x_2^T(n), \dots, x_{J-1}^T(n)]^T$  및  $\vec{x}_{n_0}(n) = [x_1(n-n_0), \dots, x_K(n-n_0)]$  일 때의  $\vec{x}$ 에 대한

자기 상관도 행렬이다.  $C$ 는 조건 행렬이고  $\vec{F} = [1, 0, \dots, 0]$  이다. 하지만 수식 (2)를 실제로 구현하는데 있어서 자기상관 행렬의 역행렬을 구하기 위해서는 많은 연산량을 필요로 한다. 따라서, 실제 구현을 할 때는 수식 (3)을 이용하여 구현한다[11].

$$\vec{W}(n+1) = P[\vec{W}(n) - \mu y(n)\vec{x}(n)] + \vec{F} \quad (3)$$

여기서  $\mu$ 는 적응필터의 학습률,  $P = I - C(C^T C)^{-1} C^T$  이다. 기존의 LCMV 빔포밍 알고리즘에서는 학습률  $\mu$ 를 고정하여 사용하고 있는데, 잡음 및 음성구간에서 그 값을 조절함으로써 더 나은 성능을 기대할 수 있다.

### 3. 제안된 알고리즘

제안된 알고리즘의 목표는 기존의 LCMV 빔포머를 활용하여 효과적으로 잡음 제거를 하는 것이다. 즉, 목표 음성구간에서는 적응필터의 학습률을 낮추어 잡음제거에 의한 음질 왜곡을 낮추고 순수 잡음구간에서는 학습률을 높여서 최대의 잡음 제거를 얻도록 한다. 이러한 동작을 위해서 본 논문에서는 채널 간 위상차를 이용한 DOA를 측정한다. 왜냐하면 일반적으로 목표 음원과 잡음원의 위치가 다르다고 가정할 수 있기 때문이다. 그림 2는 제안된 알고리즘의 전체적인 구조를 나타내고 있다. 제 1장 서론에서 언급한 바와 같이 본 연구에서는 이동형 단말기를 위한 빔포밍 알고리즘을 구현하는 것이므로 마이크로폰의 개수는 2 개를 사용하는 것으로 한다. 그림 2에서  $x_1(n)$ 과  $x_2(n)$ 은 2-채널 입력 신호이다.  $A(z)$ 는 채널보정 필터,  $w_1(z)$ ,  $w_2(z)$ 는 LCMV의 유한응답 적응필터,  $x_{2,eq}(n)$ 은  $x_2(n)$ 의 채널보상 결과,  $X_1(k)$ 와  $X_2(k)$ 는  $x_1(n)$ 와  $x_{2,eq}(n)$ 의 discrete Fourier transform (DFT)결과이다.  $\theta$ 는 DOA 이다. 마지막으로  $\hat{s}(n)$ 은 잡음이 제거된 목표 음성신호이다. 그림 2에 나타난 제안된 알고리즘의 동작은 다음과 같다. 2-채널의 입력신호가 입력되면 채널간의 주파수 차이를 보정하고 DFT를 한 후, DOA를 분석을 하여 최적의 학습률을 구하여 그 학습률로 LCMV 빔포밍을 하여 최종적으로 잡음이 제거된 신호를 얻게 된다. 주요 블록에 대한 설명은 다음과 같다.

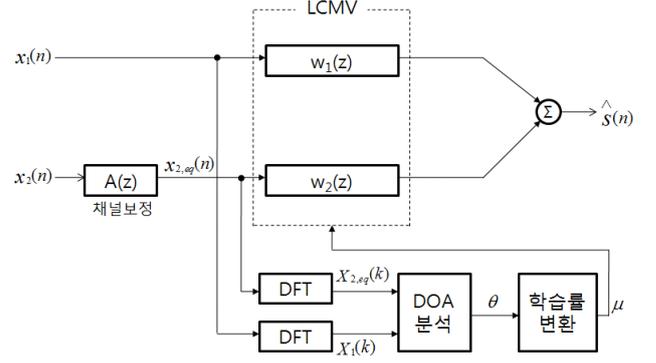


그림 2. 제안된 알고리즘의 구조

Figure 2. The structure of the proposed algorithm

#### 3.1 채널 보정

2-채널의 마이크로폰으로 신호를 받아들일 경우, 마이크로폰들의 특성 차이 때문에 두 채널 간의 스펙트럼의 차이는 피할 수 없다. 이 때, LCMV는 정면 방향으로의 신호만 유지시키기 때문에 스펙트럼의 차이가 생기면 목표 음성신호를 왜곡시킬 수 있으므로 빔포밍 성능은 저하될 수밖에 없다. 이런 문제점을 해결하기 위해서는 두 채널 간의 차이를 보상해주어야 한다. 이를 위해서 먼저 기준 채널의 스펙트럼에 다른 채널의 스펙트럼에 보상필터를 곱한 것의 차이를 수식 (4)와 같은 비용함수로 정의한다.

$$J(k) = \sum_{t=0}^{T-1} |X_1(t, k) - A(k)X_2(t, k)|^2 \quad (4)$$

여기서  $k$ ,  $t$ , 그리고  $T$ 는 각각 주파수 인덱스, 시간 인덱스, 목표 음성신호 프레임의 개수이다.  $X_1(t, k)$ 와  $X_2(t, k)$ 는 채널 1과 채널 2의 DFT 결과이다.  $A(k)$ 는 채널 보상함수의 주파수 응답이다. 비용함수를  $A(k)$ 에 대해서 최적화 하면 수식 (5)를 얻게 된다.

$$A^{(opt)}(k) = \frac{\sum_{t=0}^{T-1} X_1(t, k)X_2^*(t, k)}{\sum_{t=0}^{T-1} |X_1(t, k)|^2} \quad (5)$$

여기서  $*$ 는 복소공액 연산자이다. 최종적으로 채널 보상된 신호는 수식 (6)에 의해서 구해진다.

$$x_{2,eq}(n) = \sum_{k=0}^{L-1} a^{(opt)}(k)x_2(n-k) \quad (6)$$

여기서  $a^{(opt)}(k)$ 는 수식 (5)의 결과를 Inverse DFT 한 것,  $L$

은 채널 보상 필터의 시간 응답의 길이이다.

### 3.2 Direction of arrival 분석

최적의 학습률 변환을 설명하기에 앞서, DOA 파라미터에 대해서 설명한다. 그림 3은 DOA의 기본 개념을 나타내고 있다.

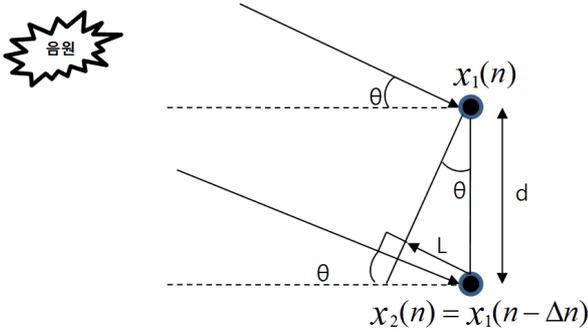


그림 3. DOA 개념  
Figure 3. The concept of DOA

그림 3에서  $d, L, \theta$ 는 마이크론의 간격, 잡음이 기준 채널과 다른 채널에 도달했을 때의 지연거리 그리고 DOA이다. DOA 유도 과정은 다음과 같다.

$$L = d \cdot \sin\theta \tag{7}$$

$$\Delta t = \frac{L}{v_s} = \frac{d \sin\theta}{v_s} \tag{8}$$

여기서  $v_s$ 는 음속이다. 표본화율을  $F_s$ 라고 할 때  $\Delta n = F_s \times \Delta t$  된다. 이를 이용하여 수식 (9)의 유도가 가능하다.

$$\Delta n = F_s \Delta t = \frac{d F_s}{v_s} \sin\theta \tag{9}$$

$x_1(n)$ 의 DFT 응답을  $X_1(k)$ 라 할 때,  $x_2(n)$ 의 DFT 응답은  $X_1(k)e^{-j\frac{2\pi}{N}k\Delta n}$  이므로, 두 응답의 비를 구하면 수식 (10)과 같다. 여기서  $N$ 은 DFT의 크기이다.

$$R = \frac{X_1(k)}{X_2(k)} = e^{j\frac{2\pi}{N}k\Delta n} \tag{10}$$

$R$ 의 위상은 수식 (13)과 같다.

$$\Phi_k = \frac{2\pi}{N}k\Delta n = \tan^{-1}\left(\frac{Im[R]}{Re[R]}\right) \tag{11}$$

여기서,  $Im[\cdot], Re[\cdot]$ 는 복소수의 허수부 및 실수부를 추출하는 연산자이다. 수식 (11)에 수식 (9)를 대입하면 수식 (12)를 얻게 된다.

$$\Phi_k(k) = \frac{2\pi}{N}k\Delta n = \frac{2\pi}{N}k\frac{dF_s}{v_s}\sin\theta_k \tag{12}$$

여기서  $\theta_k$ 는 DFT 주파수  $k$ 에서의 DOA를 나타낸다. 양변을  $\theta_k$ 로 정리하면 수식 (13)와 같다.

$$\theta_k = \sin^{-1}\left(\Phi_k \frac{Nv_s}{2\pi kdF_s}\right) \tag{13}$$

### 3.3 LCMV 필터의 적응적 학습률 변환

기존의 LCMV 빔포밍에서, LCMV 필터의 학습률은 잡음 제거를 수행하는 동안 고정된다. 학습률이 고정되어 있으면 잡음 구간과 목표 음성신호구간에 상관없이 잡음 제거를 하기 때문에 목표 음성신호구간에서 과도한 잡음 제거를 수행할 수 있으며, 잡음 구간에서는 최대로 잡음 제거를 수행할 수 없게 된다. 이러한 문제점을 해결하기 위하여 제안된 알고리즘에서는 학습률 변환 함수를 사용하여 목표신호 구간에서는 잡음제거 후의 음질을 최대로 하는 최적의 학습률을 할당하고 잡음 구간에서는 높은 학습률을 할당하여 잡음을 최대로 제거할 수 있도록 한다. 이를 위해서 DOA 기반의 학습률 변환 함수를 고안하였다. 본 논문에서의 학습률 변환 함수는 3.2절에서 설명한 DOA를 이용하여 목표 음성신호가 존재할 우도 함수를 기반으로 한다. 수식 (14)에 우도 함수를 표현하였다.

$$q_t = \frac{2}{N} \sum_{k=1}^{N/2} \left( \frac{2}{\pi} |\theta_{t,k}| - 1 \right)^2 \tag{14}$$

여기서,  $\theta_{t,k}$ 는  $t$  번째 단구간 입력에서  $k$  번째 DFT 주파수에서의 DOA이다. 수식 (14)에서  $q_t$ 는 0에서 1 사이의 값을 가지며 목표 음성구간의 값이 배경잡음 구간보다 상대적으로 큰 값을 갖게 된다. 수식 (14)의 우도 함수의 값을 지수함수를 이용한 변환을 통해서 수식 (15)의 학습률을 얻게 된다.

$$\mu_t = \begin{cases} \mu_{\max} e^{-\Delta(q_t - q_{TH})} & , q_t \geq q_{TH} \\ \mu_{\max} & , otherwise \end{cases} \tag{15}$$

이 때, 학습률이 너무 낮아지면 빔포밍에 의한 잡음제거가 이루어지지 않으므로  $\mu_t < \mu_{\min}$  이면  $\mu_t = \mu_{\min}$  을 사용한다. 그림 4는  $\mu_{\max} = 2.0$ ,  $\mu_{\min} = 0.01$ ,  $q_{TH} = 0.1$ ,  $\Delta = 30$  일 때의 학습률 변환함수를 나타낸다.

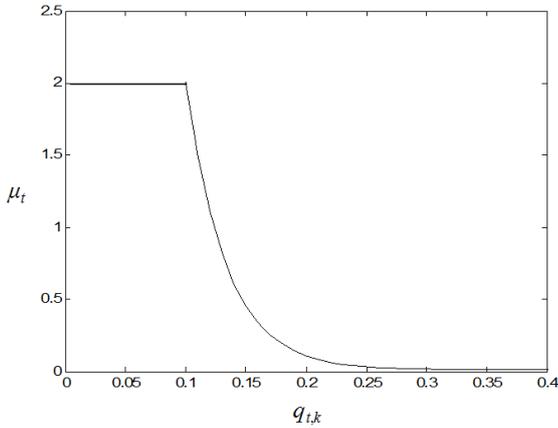


그림 4. 학습률 변환함수의 예

Figure 4. An example of learning rate mapping function

## 4. 실험 및 결과

### 4.1 DB 구성

실험을 위해 사용된 모든 데이터베이스는 사무실 환경에서 스피커로 재생되는 신호를 SONY사의 PCM-D50 모델의 녹음기를 사용하여 간격 2 cm인 2-채널 마이크로폰으로 취득되었다. 사용된 표본화율은 16 kHz 였다. 목표 음성신호는 스피커 정면 30 cm 거리에서 재생된 한국인 남녀 각 20 명이 발성한 총 60 문장이며 문장당 지속시간은 2 ~ 5 초 정도이다. 잡음 신호는 스피커 기준 60 도 방향 1.5 m 거리에서 재생된 그룹 'West Life'의 곡 'Mandy'이다. 그림 5는 DB 수집 환경을 나타내었다. 본 연구에서는 휴대단말기를 이용한 음성통화 환경에서의 성능 측정이 목표이므로 목표 음성신호원이 잡음원보다 마이크로폰 배열에 훨씬 가까이 위치하도록 하였다. SNR에 따른 성능의 측정을 위해서 목표 음성신호와 잡음 신호가 각각 녹음되며 두 신호를 인공적으로 가산하여 원하는 SNR을 맞춘다. 본 논문에서 고려된 입력 신호의 SNR은 0, 5, 10, 15, 20 dB 이다.

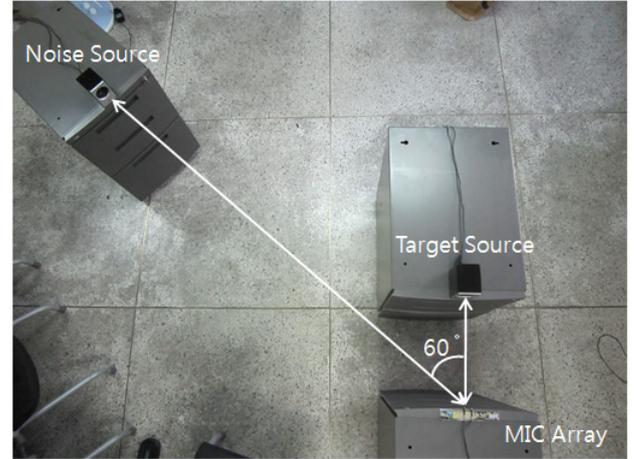


그림 5. DB 수집 환경

Figure 5. Environment for DB collection

### 4.2 성능평가

제안된 알고리즘의 성능평가를 위하여, BSD를 성능지수로 하여 SNR별로 결과를 도출하였다. 음질 측정을 위한 BSD의 기본적인 개념을 수식 (16)에 나타내었다.

$$BSD = \frac{\sum_{t=0}^{T-1} \sum_{k=0}^{M-1} (B_{t,k}^{ref} - B_{t,k})^2}{\sum_{t=0}^{T-1} \sum_{k=0}^{M-1} (B_{t,k}^{ref})^2} \quad (16)$$

여기서,  $B_{t,k}^{(ref)}$ ,  $B_{t,k}$ 는 기준신호 및 테스트 신호에 대해서 구해지는  $t$  번째 음성구간에서의  $k$  번째 Bark's scale 필터 बैं크 출력값이다.  $T$ ,  $M$ 은 단구간 음성 프레임의 개수 및 Bark's scale 필터 बैं크의 수 이다. 제안된 알고리즘에서는 총 33개의 Bark's scale 필터 बैं크가 사용되었으며 이를 위해서 10 ms 간격으로 512-fast Fourier transform (FFT)를 수행하였다.

### 4.3 파라미터 최적화

수식 (15)의 학습률 변환 함수에서 지수함수의 경사도와 밀접하게 관련있는  $\Delta$ 는 실험에 의한 전수조사 기법에 수행된다. 이 때,  $\mu_{\max}$ 는 LCMV의 적응필터가 발산하지 않는 조건에서의 최대값으로 정해져야 한다. 이 값은 실험적으로 결정되었으며 2.0 이 할당되었다.  $\mu_{\min}$  및  $q_{TH}$ 에는 실험적으로 각각 0.01 및 0.1 이 할당되었다. 채널별 적응필터의 길이는 127 이었다.

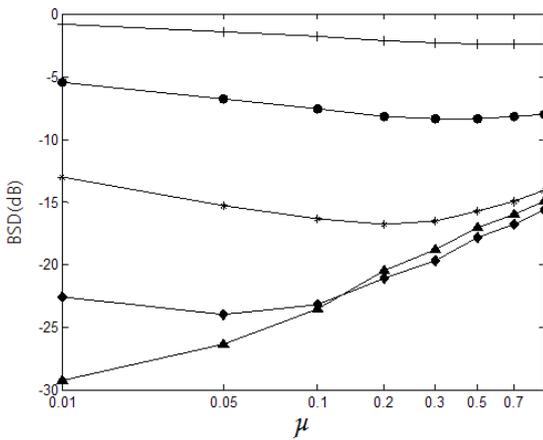


그림 6. 학습률에 따른 기존 LCMV의 성능 ('+': 입력 SNR 0 dB, '●': 5 dB, '\*': 10 dB, '◆': 15 dB, '▲': 20 dB)  
 Figure 6. Performances of the conventional LCMV beamformer according to learning rates ('+': input SNR 0 dB, '●': 5 dB, '\*': 10 dB, '◆': 15 dB, '▲': 20 dB)

4.4 실험 결과

그림 6은 기존의 고정된 적응필터의 학습률을 사용하였을 때의 LCMV의 빔포밍 알고리즘의 성능이다. 그림 6의 그래프를 보면 SNR별로 최적의 학습률이 결정되는 것을 볼 수 있다. 즉, 잡음이 클 때는 높은 학습률로 빔포밍을 해야 하고, 반대로 잡음이 작을 때는 낮은 학습률로 빔포밍을 해야 한다는 것을 나타내고 있다. 그리고 각 학습률에서 모든 SNR에 대한 평균 BSD를 성능지수로 했을 때 최적의 학습률은 0.05임을 알 수 있었다. 그림 7에서 수식 (15)의 지수함수 경사도  $\Delta$ 에 따른 전체 DB의 BSD 평균을 나타내었다. 그림 7에서 최적의 경사도는 30임을 알 수 있으며 최적치에서 기존 LCMV에 비해서 약 1.5 dB 정도의 성능 향상을 얻을 수 있음을 알 수 있다. 즉, 제안된 알고리즘에 의해서 음성구간에서 최적화된 학습률이 할당된 결과이다. 그림 8에서 잡음제거 결과에 대한 실제 음성 파형을 나타내었다. 그림 8에서 알 수 있듯이 제안된 알고리즘에 의한 학습률 할당에 의해서 잡음 구간의 전력이 기존의 알고리즘에 의한 결과와 비교했을 때 더 낮음을 알 수 있으며, 반면에, 목표 음성구간에서는 그 전력을 더 높이는 결과를 얻음을 알 수 있다. 실험결과 입력 SNR이 0, 5, 10, 15, 20 dB 일 때, 제안된 방식에 의한 평균 출력 SNR은 각각 6.2, 10.3, 14.5, 18.1, 22.0 dB 였고,  $\mu = 0.05$  를 사용한 기존 LCMV 기법으로는 각각 3.0, 7.4, 11.7, 16.1, 20.5 dB 였다.

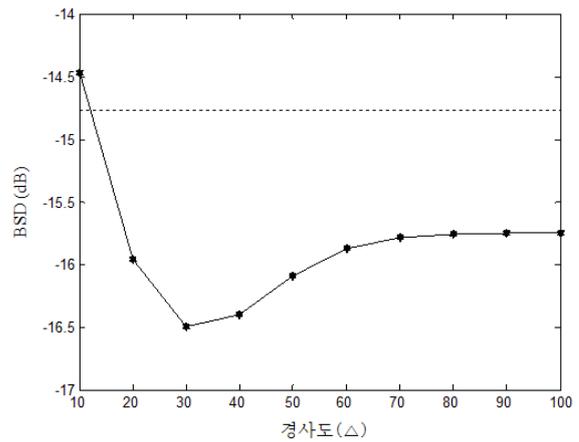


그림 7. 지수함수의 경사도에 따른 성능 변화(파선:  $\mu = 0.05$ 일 때 기존 LCMV의 성능)  
 Figure 7. Performance variation according to the slope in (15) (dashed: performance of the conventional LCMV when  $\mu = 0.05$ )

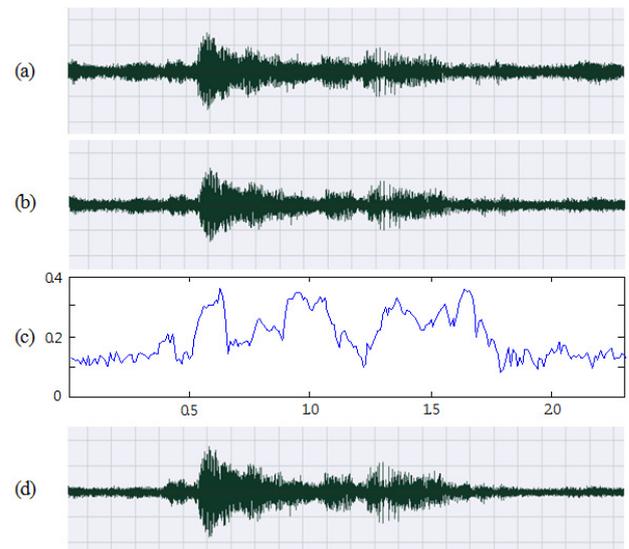


그림 8. LCMV의 빔포밍 결과들(수평축 단위: 초); (a) 잡음이 섞인 입력신호 (/내일 날씨정보 알려줘/, 입력 SNR 5 dB) (b) 기존의 LCMV 결과 ( $\mu = 0.05$ ) (d) 음성 존재 우도 (c) 제안된 LCMV 결과  
 Figure 8. Beamforming results by LCMVs (units for horizontal axis: second); (a) Noisy input waveform(/내일 날씨정보 알려줘/, input SNR 5 dB) (b) Conventional LCMV result ( $\mu = 0.05$ ) (c) Speech presence likelihood (d) Proposed LCMV result

## 5. 결론

본 논문에서는 DOA 기반 동적 학습률 조절을 이용한 다채널 음성개선 알고리즘을 제안하였다. DOA 정보로부터 학습률을 변환하기 위하여 지수함수를 응용하였고 함수의 경사도 파라미터는 전수조사 방식을 사용하여 최적화 되었다. 실험결과 기존의 LCMV와 비교했을 때, 제안된 알고리즘에 의해서 BSD 기준으로 평균 1.5 dB 정도의 성능 향상을 얻을 수 있었다. 또한, 빔포밍 후의 출력 파형에서도 배경 잡음을 더욱 효과적으로 제거할 수 있음을 알 수 있었다. 향후 연구로 빔포밍 후의 후처리를 이용한 음질 향상과 최종 잡음제거 결과에 대해서 음성 검출을 시행하여 고립 단어 수준의 음성인식률을 측정해 볼 예정이다.

## 감사의 글

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2011-0011800).

## 참고문헌

- [1] Jeong, S. and Hahn, M. (2001). "Speech quality and recognition rate improvement in car noise environments", *Electronics Letters*, Vol. 37, No. 12, pp. 801-802.
- [2] ES 202 212 V1.1.2 (2005). "Speech processing, transmission and quality aspects(STQ); distributed speech recognition; extended advanced front-end feature extraction algorithm; compression algorithm; back-end speech reconstruction algorithm", *ETSI Standard*.
- [3] Van Veen, B. D. and Buckley, K. M. (1998). "Beamforming: A versatile approach to spatial filtering", *IEEE ASSP Magazine*, Vol. 5, No. 2, pp. 4-24
- [4] Brandstein, M. and Ward, D. (2001). *Microphone Arrays: Signal Processing Techniques and Applications*, Springer.
- [5] Benesty, J., Chen, J., and Huang, Y. (2008). *Microphone Array Signal Processing (Springer Topics in Signal Processing)*, Springer.
- [6] Hyvarinen, A. and Oja, E. (2000). "Independent component analysis: Algorithms and applications", *Neural Networks*, Vol. 13, No. 4, pp. 411-430.
- [7] Lee, Y., Kim, S., Han, S., Hahn, M., Kim, Y., and Jeong, S. (2009), "Probabilistic target speech detection and its application to multi-input-based speech enhancement" *Phonetics and Speech Sciences*, Vol. 1, No. 3, pp. 97-104.

(이영재, 김수환, 한승호, 한민수, 김영일, 정상배 (2009). "확률적 목표 음성 검출을 통한 다채널 입력 기반 음성개선", *말소리와 음성과학*, 1권, 3호, pp. 97-104.)

- [8] Park, J., Lee, S., Hong, J., Jeong, S., and Hahn, M. (2008), "Beamforming optimization using filterbank-based Frost algorithm", *MALSORI*, No. 66, pp. 73-86.  
(박지훈, 이성주, 홍정표, 정상배, 한민수 (2008). "필터뱅크 기반 프로스트 알고리즘을 이용한 빔포밍 최적화", *말소리*, 66호, pp. 73-86.)
- [9] Hong, J., Han, S., Jeong, S., and Hahn, M. (2011). "Adaptive microphone array processing for high-performance speech recognition in car environment", *IEEE Transactions on Consumer Electronics*, Vol. 57, No. 1, pp. 260-266.
- [10] Wang, L., H. Ding and Yin, F. (2010), "Combining superdirective beamforming and frequency-domain blind source separation for highly reverberant signals", *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2010, pp. 1-13.
- [11] Frost, O. L. (1972). "An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, Vol. 60, No. 8, pp. 926-935.

### • 김수환 (Kim, Suhwan)

경상대학교 공과대학  
경남 진주시 가좌동 900번지  
Tel: 055-772-1727 Fax: 055-772-1729  
Email: edps2166@gnu.ac.kr  
관심분야: 음성신호처리  
현재: 경상대학교 전자공학과 석사과정

### • 이영재 (Lee, Youngjae)

경상대학교 공과대학  
경남 진주시 가좌동 900번지  
Tel: 055-772-1727 Fax: 055-772-1729  
Email: clever1999@gnu.ac.kr  
관심분야: 음성신호처리  
현재: 경상대학교 전자공학과 석사과정

### • 김영일 (Kim, Young-Il)

경상대학교 공과대학(공학연구원)  
경남 진주시 가좌동 900번지  
Tel: 055-772-1722 Fax: 055-772-1729  
Email: yi@gnu.ac.kr  
관심분야: 음향공학, 음성신호처리  
현재: 경상대학교 전자공학과 교수

### • 정상배 (Jeong, Sangbae), 교신저자

경상대학교 공과대학(공학연구원)  
경남 진주시 가좌동 900번지  
Tel: 055-772-1727 Fax: 055-772-1729

Email: [jeongsb@gnu.ac.kr](mailto:jeongsb@gnu.ac.kr)

관심분야: 음성신호처리

현재: 경상대학교 전자공학과 조교수