

## MUSIC 스펙트럼을 이용한 잡음환경에서의 목표 신호 구간 검출

### Target signal detection using MUSIC spectrum in noise environments

박 상 준<sup>1)</sup> · 정 상 배<sup>2)</sup>

Park, Sangjun · Jeong, Sangbae

#### ABSTRACT

In this paper, a target signal detection method using multiple signal classification (MUSIC) algorithm is proposed. The MUSIC algorithm is a subspace-based direction of arrival (DOA) estimation method. Using the inverse of the eigenvalue-weighted eigen spectra, the algorithm detects the DOAs of multiple sources. To apply the algorithm in target signal detection for GSC-based beamforming, we utilize its spectral response for the DOA of the target source in noisy conditions. The performance of the proposed target signal detection method is compared with those of the normalized cross-correlation (NCC), the fixed beamforming, and the power ratio method. Experimental results show that the proposed algorithm significantly outperforms the conventional ones in receiver operating characteristics (ROC) curves.

**Keywords:** target signal detection, multiple signal classification, direction of arrival, beamforming

#### 1. 서론

음성은 인간이 사용할 수 있는 가장 편리한 의사소통수단이기 때문에 음성으로 기계를 조작하는 음성인터페이스는 널리 연구되고 있다. 음성인터페이스의 성능 향상으로 최근, 스마트폰, 스마트 TV, 네비게이션 등 많은 시스템에서 음성인터페이스가 사용되고 있다. 하지만, 음성인터페이스의 성능은 주변 잡음에 의해 크게 저하되며 음성인터페이스의 성능 향상을 위해 잡음제거기술은 필수적이다.

잡음의 종류에는 정상적 잡음과 비정상적 잡음이 있다. 잡음의 특성이 시간에 따라 천천히 변하는 정상적 잡음 환경에서는 위너필터, 칼만필터 등의 단일채널 잡음 제거 기술을 적용하여 잡음을 제거할 수 있다[1][2]. 하지만 일상생활에서 대부분의 잡음은 잡음특성이 급격히 변하는 비정상적 잡음이며, 이를 제거하기 위해선 2개 이상의 마이크로폰을 사용하는 다

채널 기반 잡음 제거 기술이 필요하다. 그 중에서 Generalized Sidelobe Canceller (GSC) 기반의 빔포밍은 목적 신호의 위치 정보를 사용하기 때문에 널리 사용되고 있다[3]. GSC의 성능에 가장 영향을 주는 부분은 목표 신호 검출기인데 검출의 성능이 저하될 경우 순수 잡음 구간에서의 충분한 잡음 제거가 수행되지 못하거나 음성 구간에서의 신호 왜곡을 발생시킬 수 있다[4][5].

본 논문에서는 목표 신호 검출기 성능 향상을 위해 MUSIC 스펙트럼의 궤적을 활용한 목표 신호 검출 방법을 제안한다. MUSIC 알고리즘은 복수개의 신호원이 존재할 때 부분 공간 분석을 통한 입사각 추정 알고리즘이다[6]. MUSIC 스펙트럼은 다채널 입력 신호의 공분산 행렬을 고유값 분석함으로써 얻을 수 있으며, 추가적인 성능향상을 위해 결과 스펙트럼에 고유값의 역수로 가중치를 주었다. 이렇게 구해진 MUSIC 스펙트럼 값이 임계치 이하이면 잡음 구간으로, 이상이면 음성 구간으로 판정한다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 관련 연구들을 살펴보고 3장에서 제안된 알고리즘을 소개한다. 4장에서는 실험 결과를 분석하고 마지막으로 5장에서 본 논문의 결론을 맺는다.

1) 한국과학기술원 전기 및 전자공학과, psj@kaist.ac.kr

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

접수일자: 2012년 8월 3일

수정일자: 2012년 9월 17일

게재결정: 2012년 9월 17일

## 2. 기존 연구

### 2.1 GSC 기반 빔포밍 알고리즘

GSC 기반 빔포밍 알고리즘의 기본 구조는 <그림 1>과 같다. GSC 알고리즘은 크게 고정 빔포밍(FBF: Fixed Beamforming)단, 차단 행렬(BM: Blocking Matrix)단, 잡음제거(NC: Noise Cancellation)단으로 나눌 수 있다. 먼저, 총  $M$ 개의 다채널 입력 신호가 있을 때, 목표 음원의 위치정보(DOA: Direction Of Arrival)를 이용하여 각  $i$ 번째 채널 입력  $x_i(n)$ 의 시간 지연을 보정한다. 각 채널의 시간 지연을 보정한 후  $M$ 개의 신호를 더해지면 목표 신호는 위상정합되어 신호가 증폭되며, 목표 위치 이외의 신호는 위상부정합되어 신호가 감쇄된다. 이렇게 잡음을 제거하는 방법을 고정 빔포밍 또는 지연-합 빔포밍(Delay and sum beamforming)이라 한다. FBF단을 통과한 신호  $y_{FBF}(n)$ 에는 목표 신호가 증폭되어 있지만 잡음을 완벽히 제거할 수 없기 때문에 BM단과 NC단에서 추가적인 잡음제거를 수행한다. BM단에서는 미리 정해진 기준 채널을  $x_i(n)$ 에서 차감하여 목표 신호가 제거된 잡음 참조 신호  $v_i(n)$ 를 생성한다. NC단은 BM단에서 생성한 잡음 참조 신호  $v_i(n)$ 를 사용하여  $y_{FBF}(n)$ 에 남아있는 잔존 잡음  $v(n)$ 을 추정한다. 추정한 잔존 잡음  $v(n)$ 의 오차  $e(n)$ 이 최소화되도록 적응필터 계수를 갱신시키며, 최종적으로 출력되는 신호는  $y_{FBF}(n)$ 에서  $v(n)$ 이 제거된  $y_{GSC}(n)$ 이다. NC단의 적응필터 사용으로 더 많은 잡음을 제거할 수 있는 반면에 적응 필터계수 갱신이 목표 신호구간에서도 일어나게 되면 목표 음성구간에 심각한 신호 왜곡이 일어날 수 있다 [4][5]. 그 이유는 음원과 각 마이크로폰으로의 경로차 또는 각 마이크로폰의 특성 차이 때문에 BM단에서 목표 신호가 제거된 순수 잡음만을 얻을 수 없기 때문이다. 또한, 잡음 참조 신호  $v_i(n)$ 과 목표 신호가 상관관계에 있는 경우 왜곡이 심하게 발생한다. 특히, 잡음 신호가 인간의 음성이나 음악 신호일 때 높은 상관도를 보이며, 이 때 NC단에서 적응필터의 계수를 갱신하여 잡음 제거를 수행할 경우 목표 신호에 심각한 왜곡이 발생한다. 이러한 문제점을 해결하기 위해 사용하는 것이 적응 모드 제어기(AMC: Adaptation Mode Controller)이다. 적응 모드 제어기는 현재 입력 신호가 목표 신호 구간인지 잡음 신호 구간인지 판별하여 NC단에서 적응 필터 계수 갱신 여부를 결정한다[5]. 잡음 신호 구간으로 판별하면 필터 계수를 갱신하며 잡음제거를 수행하고, 목표 신호 구간으로 판별하면 필터 계수를 갱신하지 않아 목표 신호의 왜곡을 최소화한다. 목표 신호 구간에서 계수가 갱신되지 않으면 잡음

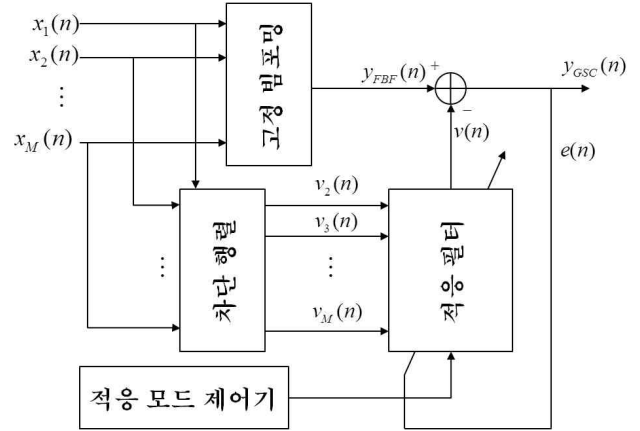


그림 1. GSC 기반 빔포밍 알고리즘의 기본 구조.

Figure 1. The basic structure of the GSC-based beamforming algorithm

또한 제거되는 양이 감소하지만 목표 신호의 왜곡을 방지할 수 있어 향상된 성능을 기대할 수 있다[4]. 적응 모드 제어기는 기존에 여러 방법이 연구되었는데, 그 중에서 널리 쓰이는 방법인 두 채널의 정규화된 상호상관도(NCC: normalized cross-correlation)를 이용하는 방법, 고정빔포밍 출력  $y_{FBF}(n)$ 의 에너지제적을 이용하는 방법, 그리고 고정빔포밍 출력  $y_{FBF}(n)$ 과 BM단 출력신호  $v_i(n)$ 의 전력비를 이용하는 방법을 소개한다.

### 2.2 종례의 적응 모드 제어 알고리즘

#### 2.2.1 두 채널의 정규화된 상호상관도를 이용하는 방법

목표 신호 구간에서 목표 음원의 DOA로 시간지연을 보상 해주면, 목표 신호가 위상정합되기 때문에 채널간의 상호상관도가 높아지게 된다. 미리 선택된 두 채널간의 정규화된 상호상관도(NCC: nomarlized cross-correlation)를 구하면 목표구간에서 높은 값, 잡음 구간에서 낮은 값을 나타낸다.  $i$ 번째 마이크로폰과  $j$ 번째 마이크로폰간의 NCC는 식 (1)에서 표현하였다.

$$P_{NCC} = \frac{\sum_{k=0}^{N-1} x_i(k - \Delta n_i) x_j(k - \Delta n_j)}{\sqrt{\sum_{k=0}^{N-1} x_i^2(k - \Delta n_i) \sum_{k=0}^{N-1} x_j^2(k - \Delta n_j)}} \quad (1)$$

식 (1)에서  $N$ 은 프레임크기를 뜻하고,  $\Delta n_i$ 은 고정 빔포밍을 위한  $i$ 번째 채널의 시간 정규화 값이다.

#### 2.2.2 고정빔포밍 출력의 에너지제적을 이용하는 방법

고정빔포밍은 목표 신호의 DOA 정보를 이용해서 그 방향

으로의 에너지가 최대가 되도록 해주는 역할을 한다. 따라서 목표 신호 구간에서는 신호가 증폭되어 원래 입력된 신호보다 개선된 신호 대 잡음비(SNR)을 갖게 되고 적절한 임계치를 사용하여 목표 구간을 검출할 수 있다[4]. 고정빔포밍 출력을 이용한 에너지제적은 식 (2)와 같다.

$$P_{FBF} = \frac{1}{N} \sum_{k=1}^N |y_{FBF}(k)|^2 \quad (2)$$

$$y_{FBF}(n) = \frac{1}{M} \sum_{m=1}^M x_m(n - \Delta n_m) \quad (3)$$

식 (3)의  $M$ ,  $\Delta n_m$ 은 마이크로폰 개수,  $m$ 번째 채널의 지연 샘플수를 뜻한다.

2.2.3 고정빔포밍 출력과 차단행렬 출력의 전력비를 이용하는 방법

GSC에서는 참조 잡음 신호를 추정하기 위해 차단 행렬 (BM : Blocking Matrix)을 사용한다[3][5]. BM단에서는 목표 음성 신호방향으로 정렬된 다채널 입력 신호에서 미리 선택된 기준 채널의 신호를 차감한다. 그 결과, GSC의 BM단에서는 목표 음성 신호가 제거된 잡음 성분의 신호만이 추출된다. 이때, FBF단 출력과 BM단 출력의 전력비를 활용하면 목표 음성 신호 구간을 검출할 수 있는데, 목표 구간에서는 BM단 출력신호에 목표 신호 성분이 사라지게 되므로 그 비가 클 것이고, 잡음 구간에서는 그렇지 않아 비가 적을 것임을 예상할 수 있다. 그러므로 전력비가 적절한 임계치보다 크면 목표구간으로, 그렇지 않으면 잡음구간으로 판별할 수 있다.  $i$ 번째 채널이 기준 채널이라 했을 때, FBF단 출력과 BM단 출력의 전력비는 식 (4)와 같다.

$$P_{PR} = \frac{\sum_{k=0}^{N-1} |y_{FBF}(k)|^2}{\sum_{k=0}^{N-1} \frac{1}{M-1} \sum_{m \neq i} |\hat{v}_m(k)|^2} \quad (4)$$

$$v_m(n) = x_m(n - \Delta n_m) - x_i(n - \Delta n_i) \quad (5)$$

식 (5)의  $v_m(n)$ 은  $m$ 번째 채널의 BM단 출력 신호이다.

### 2.3. MUSIC 알고리즘

다채널 신호의 상호 상관 행렬을 구한 후, 이 행렬을 고유 값 분해하면 신호 부분 공간과 잡음 부분 공간으로 분리할 수

있다. 이렇게 구해진 부분공간을 이용하여 DOA를 추정하는 방법을 부분공간 기반 DOA 추정기법, 또는 Multiple Signal Classification(MUSIC) 알고리즘이라 한다. MUSIC 알고리즘은 Schmidt에 의해 제안된 알고리즘으로 협대역 신호의 DOA 추정에 널리 쓰이고 있는 방법 중 하나이다[6].  $M$ 개의 채널에  $D$ 개의 신호가 입력되고 있을 때,  $M$ 개의 채널 입력 신호 벡터  $X$ 는 식 (6), (7)과 같이 나타낼 수 있다.

$$\vec{X} = A\vec{F} + \vec{W} \quad (6)$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} \vec{a}(\theta_1) & \vec{a}(\theta_2) & \cdots & \vec{a}(\theta_D) \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_D \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_M \end{bmatrix} \quad (7)$$

벡터  $X$ ,  $F$ 는 모두 복소값으로 표현된다. 벡터  $\vec{F}$ 는  $D$ 개의 음원에서 발생하는 신호이며, 벡터  $\vec{W}$ 는 각 마이크로폰에서 발생하는 가산성 잡음이다. 벡터  $\vec{a}(\theta_i)$ 는 이득과 위상차를 보정해주는 함수이며 마이크로폰 배치와 음원의 입력 경로 등에 따라 결정된다. 벡터  $\vec{F}$ 가  $w_c$ 를 중간 주파수(center frequency)로 가지는 협대역 신호이고 마이크로폰이 선형으로 배치되어 있을 때, 벡터  $\vec{a}(\theta)$ 는 식 (8)과 같다.

$$\vec{a}(\theta) = \begin{bmatrix} a_1 e^{-jw_c \Delta t_1} \\ a_2 e^{-jw_c \Delta t_2} \\ \vdots \\ a_M e^{-jw_c \Delta t_M} \end{bmatrix} \quad (8)$$

식 (8)에서  $\theta$ ,  $a_i$ ,  $\Delta t_i$ 는 각각 DOA, 각 채널의 이득, 각 채널의 시간 지연을 뜻한다. 행렬  $X$ 의 공분산 행렬은 식 (9)로부터 얻을 수 있다.

$$R = E[\vec{X}\vec{X}^H] \quad (9)$$

식 (9)에서  $E[\cdot]$ 는 통계적 평균 연산자이며  $^H$ 는 허미시안 전치(Hermitian transpose) 연산자이다.  $R$ 의 고유값 분해를 통해  $M$ 차원의 행렬을  $D$ 차원의 신호 부분공간과  $(M-D)$ 차원의 잡음 부분공간으로 분리할 수 있다[6].  $R$ 의 고유값 중 가장 높은  $D$ 개의 값에 해당하는 고유벡터를 확장시켜 만든 부분공간을 신호 부분공간  $E_S$ , 나머지  $M-D$ 개의 고유벡터를 확장시켜 만든 부분공간을 잡음 부분공간  $E_N$ 일 때 식

(10), (11)처럼 표현할 수 있다.

$$E_N = [\vec{e}_1 \ \vec{e}_2 \ \cdots \ \vec{e}_{M-D}] \quad (10)$$

$$E_S = [\vec{e}_{M-D+1} \ \vec{e}_{M-D+2} \ \cdots \ \vec{e}_M] \quad (11)$$

(단,  $\lambda_1 < \cdots < \lambda_{M-D} \ll \lambda_{M-D+1} < \cdots < \lambda_M$ )

여기서  $\vec{e}_i$ 는  $i$ 번째 고유벡터이며,  $\lambda_i$ 는  $\vec{e}_i$ 에 해당하는 고유값이다. MUSIC 알고리즘은  $\theta$  방향에서 신호가 올 때, 벡터  $\vec{a}(\theta)$ 가 신호 부분공간  $E_S$ 와 가까워진다는 사실을 이용한다 [6][7]. 신호 부분공간  $E_S$ 로부터  $\vec{a}(\theta)$ 의 유클리디안 거리는  $\vec{a}(\theta)$ 를 잡음 부분공간  $E_N$ 에 정사영 시킨 길이와 같으므로 식 (12)와 같이 표현할 수 있고, MUSIC 스펙트럼은 식 (12)에 역수를 취해서 식 (13)처럼 얻을 수 있다.

$$|d(\theta)|^2 = \vec{a}(\theta)^H E_N E_N^H \vec{a}(\theta) = \sum_{i=1}^{M-D} |\vec{e}_i^H \vec{a}(\theta)|^2 \quad (12)$$

$$P_{MU}(\theta) = \frac{1}{|d(\theta)|^2} = \frac{1}{\sum_{i=1}^{M-D} |\vec{e}_i^H \vec{a}(\theta)|^2} \quad (13)$$

### 2.3.1 MUSIC 스펙트럼 가중치 적용

MUSIC 알고리즘은 미리 음원의 개수를 알고 있어야 한다. 고유값을 이용해 신호개수를 추정 가능하지만, 잡음이 섞여있으면 정확한 추정이 불가능하기 때문에 성능이 감소하게 된다. 또한, 고유벡터를 신호 부분공간 또는 잡음 부분공간으로 경판정(hard-decision)함으로써 고유값이 가지는 정보를 충분히 이용하지 못한다. 이러한 단점을 보완하기 위해 고유값을 이용하여 MUSIC 스펙트럼에 가중치를 적용한다. 모든 고유벡터를 사용하면서 고유값으로 가중치를 준 MUSIC 스펙트럼을 식 (14)에 표현하였다.

$$P_{MU, Weighted}(\theta) = \frac{1}{\sum_{i=1}^M \lambda_i |\vec{e}_i^H \vec{a}(\theta)|^2} \quad (14)$$

이 방법은 음원의 개수를 미리 알고 있지 않아도 적용 가능하며 목표 음원 DOA 방향의 MUSIC 스펙트럼을 크게 증폭시켜 주어 더 정확하게 DOA를 추정할 수 있다.

## 3. 제안된 적응 모드 제어 알고리즘

기존의 목표 신호 검출방법들은 다음과 같은 문제점이 존재한다. 실제 환경에서 NCC는 두 채널간의 부정합이 존재할 수 있어서 성능이 떨어지게 된다. 게다가 오직 두 채널의 정보만을 이용한다는 단점이 있다. 고정빔포밍 출력은 목표 신호가 위상정합된다 하더라도 SNR을 향상시키는 데 한계가 있고, 고정빔포밍 출력과 차단행렬 출력의 전력비를 이용하는 방법은 차단행렬을 단순 차감할 경우에 목표 신호 유출(target signal leakage)이 발생하여 목표구간에서도 전력비가 낮아질 수 있는 단점이 있다[5]. 이러한 단점들을 보완하기 위해 본 논문에서는 DOA 추정 기법인 MUSIC 알고리즘을 이용한 적응 모드 제어 방법을 제안한다. MUSIC 알고리즘은 협대역 복소 신호에 대한 DOA 추정 기법이기 때문에 빔포밍 환경에 적용하기 위하여 <그림 2>와 같은 알고리즘을 제안하였다. <그림 2>에서  $\vec{X}(n)$ 은 M개의 채널 입력 신호이다. 각 채널의 신호는 대역통과 필터를 통과하여 협대역신호인  $\vec{X}_{BP}(n)$ 로 변환되고, 해석 신호 생성기(ASG: Analytic Signal Generation)를 통과한다. 생성된 해석 신호  $\vec{X}_A(n)$ 은 MUSIC 스펙트럼 계산에 사용되며, 적응 모드 제어 성능을 향상시키기 위해 MUSIC 스펙트럼에 고유값을 이용하여 가중치를 주었다. 목표 음원 DOA가  $\theta_T$ 로 주어졌을 때, 결과 MUSIC 스펙트럼은  $P_{MU, Weighted}(\theta_T)$ 이고, 목표 신호 구간 판별기는 구해진 MUSIC 스펙트럼이 임계치  $P_{TH}$ 보다 크면 목표 구간으로 작으면 잡음 구간으로 판별한다. 각 블록에 대한 자세한 설명은 다음과 같다.

### 3.1 대역통과 필터

2.3절의 MUSIC 알고리즘의 모델에 따르면  $\vec{a}(\theta)$ 는 복소 신호에 곱해서 위상 변이(phase shift) 시키고, 각 채널의 시간 지연을 보상하게 되는데, 이처럼  $\vec{a}(\theta)$ 를 이용하여 복소 신호를 시간 지연시키기 위해선  $\vec{F}$ 는 협대역 신호여야 한다[6][7]. 협대역 신호를 생성하는 방법은 크게 두 가지이다. 하나는 이산 푸리에 변환(DFT: Discrete Fourier Transform)을 수행한 후, 강한 주파수 성분만을 추출하는 방법이고, 또 하나는

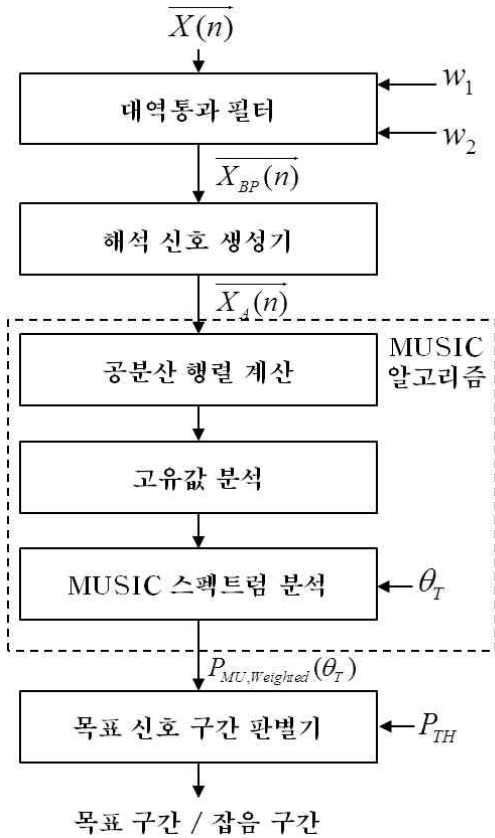


그림 2. 제안된 목표 신호 구간 검출기 블록다이어그램  
Figure 2. The block diagram of proposed target signal detection method.

좁은 대역만을 통과시키는 대역통과필터(BPF: BandPass Filter)를 적용하는 방법이다. 강한 주파수 피크 성분을 추출하는 방법은 추출한 성분이 목표 신호 성분인지 잡음 신호 성분인지 알 수 없고, 이를 보완하기 위해 강한 주파수 성분을 여러개 추출하여 같이 사용한다 하더라도 비용(cost)이 증가하는 단점이 있다. 그러므로 본 논문에서는 협대역 신호를 생성하기 위해 대역통과필터를 적용하였다. 대역 중간 주파수  $w_c$ 는 식 (15)와 같다.

$$w_c = \frac{(w_1 + w_2)}{2} \tag{15}$$

$w_c$ 는 식 (14)에서  $P_{MU, Weighted}(\theta)$  계산에 필요한  $\vec{a}(\theta)$ 를 설계하는데 사용된다. 통과대역에서 목표 신호와 잡음 신호의 중복은 피할 수 없지만 잡음 신호에 비해 목표 신호의 전력이 클수록 MUSIC 스펙트럼의 결과가 정확하므로,  $w_1, w_2$ 의 선택이 중요하다.

### 3.2 해석 신호 생성(ASG: Analytic Signal Generation)

MUSIC 알고리즘의 모델은 식 (6), (7)의 음원 신호  $\vec{F}$ 를 복소 신호로 가정하고 복소 함수  $\vec{a}(\theta)$ 를 곱해주어 시간 지연을 수행한다. 이렇게 생성된 채널 입력신호  $\vec{X}$  또한 복소 신호이기 때문에 MUSIC 알고리즘을 적용하기 위해선 <그림 2>의  $\vec{X}(n)$ 도 복소 신호이어야 한다. 하지만, 우리가 실제 환경에서 마이크로폰으로 입력받는 신호는 실수 신호이므로 이를 복소 신호로 표현해주어야 한다. 이처럼 실제 환경의 실수 신호를 이론적 접근을 위해 복소 신호로 표현하는 것을 해석 신호 생성이라 하며 표현된 복소 신호는 해석 신호 (Analytic Signal)이라고 한다[7]. 입력 신호를  $x(n)$ , 출력된 해석 신호를  $x_a(n)$ 이라 할 때, 식 (16)을 만족한다.

$$x_a(n) = \int_0^\pi 2X(w)e^{jwn} dw \tag{16}$$

식 (16)에서  $X(w)$ 는  $x(n)$ 을 이산 푸리에 변환한 결과이다.

### 3.3 목표 신호 구간 판별

목표 목표 신호 구간 판별기는 MUSIC 스펙트럼 분석을 통해 계산된  $P_{MU, Weighted}(\theta_T)$ 를 사용한다. 주어진 임계치  $P_{TH}$ 의 값과 비교하여 크면 목표 구간, 작으면 잡음 구간으로 판별한다. 임계치  $P_{TH}$ 의 값은 절대적인 값을 사용하지 않고 순수 잡음 구간으로 간주되는 초기 입력 구간에 대해 추출한  $P_{MU, Weighted}(\theta_T)$ 의 평균값에 이득을 곱해 계산한다. 식 (17)에 임계치를 계산하는 방법을 표현하였다.

$$P_{TH} = \alpha_{TH} P_{MU, Weighted}^{(mean)}(\theta_T) \tag{17}$$

식 (17)에서  $\alpha_{TH}$ 는 초기 파라미터 평균값에 곱해주는 이득이며  $P_{MU, Weighted}^{(mean)}(\theta_T)$ 는 순수 잡음 구간의 MUSIC 스펙트럼 평균값이다.

## 4. 실험 및 결과

### 4.1 다채널 DB 수집 및 실험 조건

목표 신호와 잡음 신호는 6×5×3m의 일반 사무실환경에서 각각 따로 녹음되었다. 잡음이 섞인 목표 신호를 얻기 위하여 잡음 신호 구간을 임의로 선택한 후, SNR별로 인위적으로 합

성하였다. 실험에 사용한 SNR은 0, 5, 10, 15, 25 dB이며, 시작과 끝 각 1초는 목표 신호가 존재하지 않는 순수잡음구간이다. 목표 음성 신호와 잡음 신호는 각 1개를 사용하였고 마이크로폰과 소스간 거리는 2m, 목표 음원은 정면에서, 잡음 음원은 45도 각도에서 입력된다고 가정하였다. 잡음 신호는 TV 소리로 가정하고 토크쇼 소리를 사용하였다. 목표 신호는 한국어 PBW(Phonetically Balanced Words) 452 DB 4세트를 사용하였고, 5가지 SNR로 합성되어 총  $452 \times 4 \times 5 = 9040$ 개의 샘플이 실험에 사용되었다. 잡음 신호 및 목표 신호는 모두 스피커를 통해서 재생되었다. 7개의 마이크로폰을 2cm 간격으로 선형 배치하여, 16kHz의 표본화율과 16 bit의 해상도로 녹음하였다.

<그림 2>의 각 파라미터들은 다음과 같이 설정하였다. 3.1절에서 언급된 조건에 따라  $[w_1 = 100, w_2 = 500]$  Hz의 대역을 통과 대역으로 사용하였다. 필터는 5차 엘립틱(elliptic) 필터를 설계하여 사용하였다. 목표 음원이 정면에 있다고 가정하였으므로,  $\theta_T$ 를  $0^\circ$ 로 설정하여  $P_{MU, Weighted}(\theta_T)$ 를 계산하였다.  $P_{TH}$ 는  $\alpha_{TH}$ 를 100단계로 이동시키며 식 (17)에 의해 구해지는 값 100개를 사용하였다.  $\alpha_{TH}$ 는 주어진 DB에서 가질 수 있는 최소값과 최대값을 구하여 100단계로 이동시켰다.  $P_{TH}$ 를 계산하는데 사용한 초기 순수잡음구간은 30 ~ 230ms 구간으로 설정하였다. 모든 알고리즘의 구간 판정은 제안된 알고리즘과의 비교를 위해 절대적인 값을 사용하지 않고 3.3절의 제안된 알고리즘 판별 방법을 사용하였다.

4.2 성능 평가

제안된 알고리즘의 성능은 2.2절에서 소개한 기준 목표 신호 검출 방법과 비교하여 평가되었다. 알고리즘 별 결과파형은 <그림 3>과 같다. 제안한 방식을 사용할 경우 파형이 가장 깔끔하게 나오는 것을 확인할 수 있다. 특히, 잡음구간에서 다른 방법에 비해 0에 가까운 낮은 값을 가져 성능이 뛰어날 것으로 기대된다.

결과의 타당성을 확인하기 위해 4.1절에서 언급한 방법으로 구해진 100개의  $P_{TH}$ 에 의해서 생성된 receiver operating characteristics (ROC) 곡선을 <그림 4>에 나타내었다. x축은 잡음 구간을 목표 구간으로 오판별하는 비율 false acceptance rate (FAR)을, y축은 목표 구간을 목표 구간으로 판별하는 비율 true acceptance rate (TAR)를 뜻한다. ROC 곡선을 평가하는 척도인 area under the ROC curve (AUC)는 <표 1>과 같다. ROC 곡선은 왼쪽 위로 치우칠수록 즉, AUC가 클수록 좋은 성능을 가지는 것으로 해석할 수 있다. 제안한 방법이 가장 좋은 성능을 보였으며, 그 다음으로 고정빔포밍 출력을 사용하는 방법, 기존 MUSIC 스펙트럼을 사용하는 방법

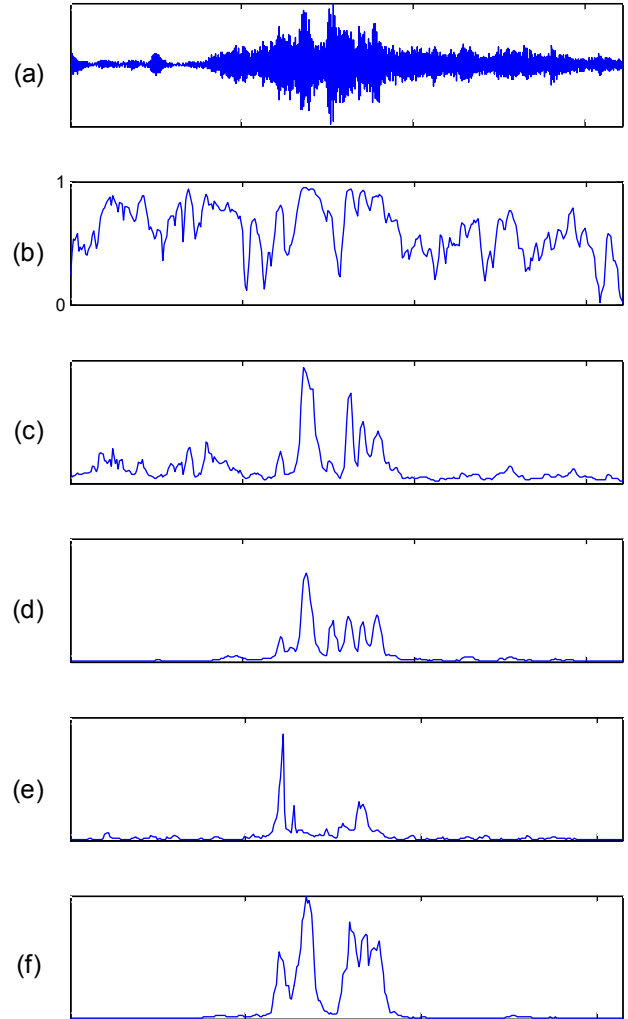


그림 3. (a) SNR 5dB 입력 신호 (b) NCC 궤적 (c) 전력비 궤적 (d) FBF 궤적 (e) MUSIC 스펙트럼 궤적 (f) Weighted MUSIC 스펙트럼 궤적 (\*수평축의 한칸은 1초를 해당함)  
Figure 3. (a) Input signal with 5dB SNR (b) Normalized Cross Correlation contour (c) Power ratio contour between  $y_{FBF}(n)$  and  $\hat{v}_v(n)$  (d) Power contour of fixed beamforming result (e) MUSIC spectrum contour (f) Weighted MUSIC spectrum contour (\*One second for a tick on the horizontal axis)

이 좋은 성능을 보였다. 두 채널간 NCC를 이용하는 방법이 성능이 가장 좋지 않았다. 목표 신호 검출기의 일반적인 기준인 TAR을 80%로 고정시키면, 제안한 방법의 FAR은 고정빔포밍 출력을 사용하는 방법보다 약 15 % 개선되었다.

입력 신호의 SNR 별 최소 오차율은 <그림 5>에 나타내었다. 오차율은 목표 구간을 잡음 구간으로 판별하는 오류와 잡음 구간을 목표 구간으로 판별하는 오류를 모두 포함한다. 최소 오차율에 사용된 이득 임계치 값은 실험적으로 결정되었다. 제안한 방법은 고정빔포밍 출력을 사용하는 방법보다 오차율이 약 5% 감소하여 13.5%의 오차율을 보였다.

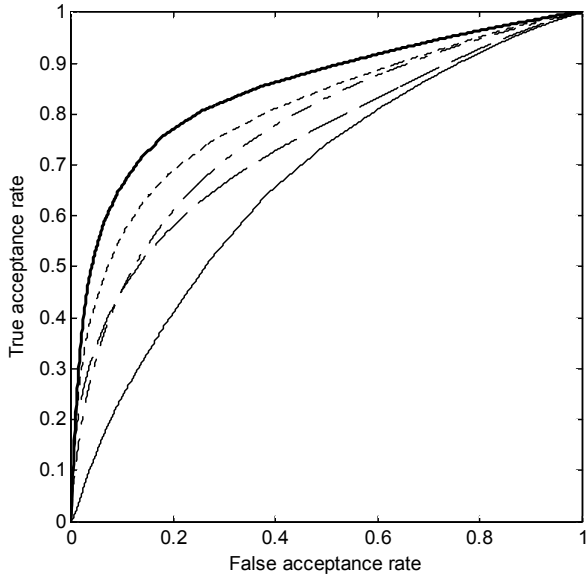


그림 4. ROC 곡선

Figure 4. ROC curves

(soild : NCC, dashed : Power Ratio, dashed-dotted: MUSIC dotted : FBF, thick soild : Weighted MUSIC )

표 1. 알고리즘별 AUC

알고리즘	AUC (Area Under the ROC Curve)
NCC	0.6685
Power Ratio	0.7431
FBF	0.8051
MUSIC	0.7690
Weighted-MUSIC	<b>0.8494</b>

### 5. 결론

본 논문에서는 MUSIC 알고리즘 기반 DOA 추정기법을 활용한 목표 신호 구간 검출 알고리즘을 제안하였다. 실험결과, 제안한 방법의 오차율은 기존 알고리즘 가장 좋은 성능을 보였던 고정빔포밍의 출력을 이용하는 방법에 비해 약 5% 더 감소하였다. 향후, 제안한 목표 신호 검출기를 GSC 구조에 결합하여 잡음제거 성능을 향상시키고 음성인식기의 끝점 검출 기에도 활용하여 MUSIC 스펙트럼 기반 다채널 음성인터페이스에 대해 연구할 예정이다.

### 감사의 글

이 논문은 지식경제부와 산업기술평가관리원의 정보통신연구개발사업[10041610, 인식센서융합 기반 실환경 하에서 임의의 사용자 30명에 대해 인식률 99%에 근접하는 사용자의 신원과 행위 및 위치 정보 인식 기술 개발 (1차년도)]의 지원을 받아 수행되었습니다.

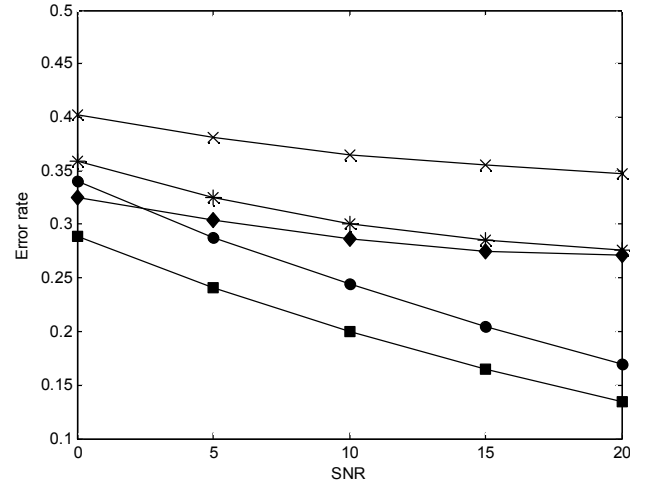


그림 5. SNR (dB) 별 최소 오차율

Figure 5. Minimum error rates for each SNR

(‘×’: NCC, ‘\*’: Power Ratio, ‘◆’: MUSIC, ‘●’: FBF, ‘■’: Weighted MUSIC)

### 참고문헌

- [1] ETSI (2005). *Speech processing, transmission & quality aspects (STQ)*, v.1.1.2, ES 202 212.
- [2] Jeong, S. & Hahn, M. (2001). Speech quality and recognition rate improvement in car noise environments, *Electronics Letters*. Vol. 37, No. 12, 801-802.
- [3] Benesty, J., Chen, J. & Huang, Y. (2008). *Microphone array signal processing (Springer topics in signal processing)*, Springer.
- [4] Lee, Y., Kim, S., Han, S., Hahn, M., Kim, Y. & Jeong, S. (2009). Probabilistic Target Speech Detection and Its Application to Multi-Input-Based Speech Enhancement, *Phonetics and Speech Sciences*. Vol. 1, No. 3, 97-104. (이영재, 김수환, 한승호, 한민수, 김영일, 정상배 (2009). 확률적 목표 음성 검출을 통한 다채널 입력 기반 음성 개선, 말소리와 음성과학. 1권, 3호, 97-104.)
- [5] Hoshuyama, O. et al. (1999). A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters, *IEEE Trans. Signal Proc.* Vol 47, No. 10, 2677-2688.
- [6] Schmidt, R. O. (1986). Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas and Propagation*. Vol 34, No. 3, 276-280.
- [7] Varma, K. (2002). Time-delay-estimate based direction-of-arrival estimation for speech in reverberant environments, M. S. thesis, Dept. of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Virginia, USA

- [8] Hayes, M. (1996). *Statistical digital signal processing and modeling*, John Wiley & Sons.
- [9] Hwang H. K, et al. (2008). Direction of arrival estimation using a root-MUSIC algorithm, *Proceedings of IMECS 2008*. Vol. 2, 19-21.

• **박상준 (Park, Sangjun)**

한국과학기술원 전기 및 전자공학과  
대전광역시 유성구 대학로 291  
Tel: 042-350-8074 Fax: 042-350-7619  
Email: psj@kaist.ac.kr  
관심분야: 음성신호처리  
현재: 한국과학기술원 전기 및 전자공학과 박사과정

• **정상배 (Jeong, Sangbae)** 교신저자

경상대학교 공과대학(공학연구원)  
경남 진주시 가좌동 900번지  
Tel: 055-772-1727 Fax: 055-772-1729  
Email: jeongsb@gnu.ac.kr  
관심분야: 음성신호처리  
현재: 경상대학교 전자공학과 조교수