

## GSC 기반 빔포밍을 위한 주파수 밴드별 전력비 분포의 혼합 가우시안 모델을 이용한 목표 음성신호의 검출

장형욱 · 김영일 · 정상배\*

### Target Speech Detection Using Gaussian Mixture Model of Frequency Bandwise Power Ratio for GSC-Based Beamforming

Hyungwook Chang · Youngil Kim · Sangbae Jeong\*

Department of Electronics Engineering, Gyeongsang National University, Jinju 660-701, Korea

#### 요 약

다양한 종류의 잡음에 의해서 발생하는 음성인식 성능 저하를 보상하기 위해서는 잡음제거가 필수적이다. 마이크로폰 배열을 이용하는 많은 잡음제거 기술 중에서, GSC는 비정상성 잡음을 제거하기 위해서 널리 적용되어 왔다. GSC의 성능은 AMC에 의해서 직접적인 영향을 받는다. 즉, 정확한 목표 음성 신호의 검출은 순수 잡음구간에서의 충분한 잡음제거 및 목표 음성구간에서의 적은 왜곡을 보장하기 위해서 필수적이다. 따라서, 본 논문에서는 고정 빔포밍의 출력과 차단 매트릭스의 출력간의 전력비가 주파수 밴드 단위로 계산되는 향상된 AMC 설계법을 제안한다. 그 후, 밴드별 전력비는 가우시안 혼합에 의해서 각 클래스가 확률적으로 모델링 된다. 실험결과, 제안한 알고리즘이 ROC 및 출력 SNR 관점에서 더 높은 성능을 보였다.

#### ABSTRACT

Noise reduction is necessary to compensate for the degradation of recognition performance by various types of noises. Among many noise reduction techniques using microphone array, generalized sidelobe canceller (GSC) has been widely applied to reduce nonstationary noises. The performance of GSC is directly affected by its adaptation mode controller (AMC). That is, accurate target speech detection is essential to guarantee the sufficient noise reduction in pure noise intervals and the less distortion in target speech intervals. Thus, this paper proposes an improved AMC design technique in which the power ratio of the output of fixed beamforming to that of blocking matrix is calculated frequency bandwise and probabilistically modeled by mixture Gaussians for each class. Experimental results show that the proposed algorithm outperforms conventional AMCs in receiver operating curves (ROC) and output SNRs.

**키워드** : 잡음제거, 마이크로폰 배열, 일반화된 부극 제거기

**Key word** : noise reduction, microphone array, generalized sidelobe canceller

접수일자 : 2014. 11. 21 심사완료일자 : 2014. 12. 08 게재확정일자 : 2014. 12. 22

\* **Corresponding Author** Sangbae Jeong(E-mail:jeongsb@gnu.ac.kr, Tel:+82-55-772-1727)

Department of Electronics Engineering/ERI, Gyeongsang National University, Jinju 660-701, Korea

**Open Access** <http://dx.doi.org/10.6109/jkiice.2015.19.1.61>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서 론

여러 가지 기계-인간 인터페이스 기술 중에서 음성은 가장 편리한 방법이며 다양한 연구가 이루어져 왔다. 현재, 스마트폰과 네비게이션, 지능형 로봇, 가전제품 등 많은 곳에서 음성 기반의 인터페이스가 적용되고 있다. 음성인터페이스에서 음성 인식 기술은 입력 음성을 분석하여 기계의 작동을 위한 명령으로 변환한다. 그런데 음성 인식 기술의 성능은 가산적인 잡음에 의하여 심하게 저하되는 경향이 있다. 따라서, 실제 환경에서의 안정적인 성능 확보를 위해서 다양한 잡음 제거 기술들이 연구되어왔다[1-5].

잡음의 종류로는 정상성 잡음과 비정상성 잡음이 있다. 정상성 잡음으로 PC의 팬 소리, 자동차 엔진 소리 등이 있으며, 비정상성 잡음으로는 사람의 목소리, 음악 소리 등이 있다. 정상성 잡음을 제거하는 방법으로 단일 채널을 이용한 Wiener나 Kalman 필터링과 같은 적응 필터링이 있고[1, 2], 비정상성 잡음은 마이크로폰 배열을 이용하여 제거하며 목표 음원에 대한 정보가 필요 없는 blind source separation (BSS) 기술과 목표 신호의 위치정보가 필요한 빔포밍 기술이 연구되어왔다[3].

대표적인 다채널기반 빔포밍 알고리즘으로는 linearly constrained minimum variance (LCMV)와 일반화된 부극제거기(generalized sidelobe canceller (GSC)) 알고리즘이 있다[4, 5]. LCMV 기반의 빔포밍 알고리즘은 목표 음원 방향의 신호 왜곡이 없는 조건에서 출력 에너지를 최소화시킨다. 이러한 조건 때문에 적응 필터의 수렴이 늦어짐으로 인한 충분한 잡음제거가 이루어지지 않는 단점이 있다[4]. GSC 빔포밍 알고리즘은 LCMV 알고리즘을 기능적으로 분해해 놓은 형태와 동등함이 알려져 있으며 adaptation mode controller (AMC)만 제대로 동작할 경우에 LCMV 보다 더 훌륭한 성능을 얻을 수 있음이 알려져 있다[5]. GSC 빔포밍에 필수적인 AMC의 역할은 잡음이 섞인 다채널 입력 신호에 대해서 그것이 순수 잡음 구간인지 아니면 목표 음성구간인지를 판정하는 것이다. 만약, 목표 음성구간이라면 GSC의 noise canceller (NC)에 포함되어 있는 적응 필터의 계수 갱신을 중지하여 잡음제거 후의 목표 음성 왜곡을 줄이게 한다. 따라서, AMC는 GSC 빔포밍의 안정적인 성능에 매우 중요한 역할을 한다.

본 논문에서는 정확한 목표신호 구간 검출을 위해서

주파수 밴드 단위에서의 fixed beamforming (FBF) 출력과 blocking matrix (BM) 출력간의 전력비를 추정하여 AMC를 설계하는 방식을 제안한다. 본 논문의 구성은 다음과 같다. 제 2장에서 GSC 기반의 빔포밍과 기존의 목표 음성 구간 검출 기법, 가우시안 혼합 모델의 이론을 소개한다. 제 3장에서는 제안된 목표 신호 검출 알고리즘을 각 단계별로 설명한다. 제 4장에서는 제안된 알고리즘 및 비교 알고리즘들의 실험결과에 대해서 논하고 마지막으로 제 5장에서 본 논문의 결론을 맺는다.

## II. 관련 연구

### 2.1. GSC 기반의 빔포밍

그림 1에서 주파수 영역에서 동작하는 GSC 기반의 빔포밍의 블록 다이어그램을 나타내었다. 먼저, 다채널 단구간 입력에 대해서 time delay compensation (TDC)가 수행되어 목표 음원 방향에 대한 채널 간 시간차가 없어지도록 한다. 그 후, 신호를 채널별로 discrete Fourier transform (DFT)를 수행하여 주파수 영역으로 변환한다. 실질적인 잡음 제거를 수행하기 전에 FBF를 통해서 목표 음성의 에너지가 가장 커지도록 하며, 임의의 채널값에서 기준 채널값을 차감함에 의해서 목표 음성 신호를 소거하여 참조 잡음 신호를 취득하기 위한 BM이 실행된다. FBF의 출력이 잔존하는 잡음과 BM의 출력에서 취득되는 참조 잡음 신호는 그 모양이 다를 수 있으므로 NC 블록의 적응 필터를 이용하여 최종적인 잡음제거가 이루어지게 된다. NC 블록의 적응 필터는 least mean square (LMS) 형태의 알고리즘에 의해서 그것의 계수가 갱신된다. 이때, 만약, 목표 음성구간에서도 NC 블록의 적응 필터 계수가 갱신된다면 잡음제거 후의 음성 왜곡이 심하게 발생하게 된다. 그 이유는 목표 음성구간에서 BM에서 목표 음성의 leakage가 발생할 수 있기 때문이다. 따라서, GSC의 이러한 문제점을 완화시키기 위해서 AMC가 필수적으로 구현되어 입력된 다채널 신호가 목표 음성구간인지 아니면 순수 잡음구간인지를 판단해야 한다. AMC의 결과를 참조하여 입력 신호가 목표 음성구간일 경우에 적응 필터의 계수 갱신을 중지시켜야 한다. 최종적으로 잡음이 제거된 신호는 inverse DFT (IDFT)에 의해서 시간 영역의 신호로 변환된다.

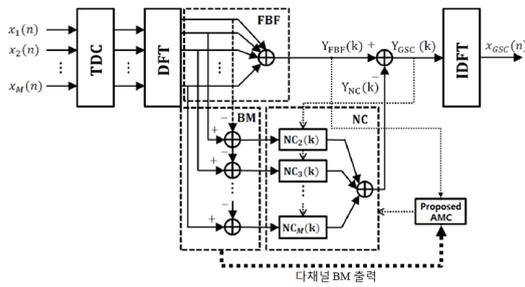


그림 1. GSC 빔포밍 블록 다이어그램  
Fig. 1 Block diagram of GSC beamforming

### 2.2. Adaptation mode controller (AMC)

GSC 기반의 빔포밍에서 NC 필터 계수는 고정 빔포밍 출력에서 목표 신호 성분을 제거하지 않기 위해 오직 잡음구간에서만 갱신되어야만 한다. 따라서 목표 신호의 왜곡이 일어나지 않기 위해서 목표 음성 구간 검출 기법이 필요하며 그 방법으로는 정규화된 상관도를 이용한 목표 신호 검출방법, 전력비를 이용한 목표 신호 검출방법 등이 있다[6, 7].

정규화된 상관도를 이용한 목표 신호 검출방법은 시간 영역에서 임의의 두 채널 간의 정규화된 상호상관도를 이용하여 목표 신호의 존재 유무를 결정하며 식 (1) 과 같이 계산된다[6].

$$\phi_t^{(NCC)} = \frac{\sum_{k=k_1}^{k_2} X_i(k) X_j^*(k)}{\sqrt{\sum_{k=k_1}^{k_2} |X_i(k)|^2} \sqrt{\sum_{k=k_1}^{k_2} |X_j(k)|^2}} \quad (1)$$

여기서,  $t$  는 단구간 프레임의 인덱스,  $(\cdot)^*$  는 복소 공액 연산자이다.  $X_i(k)$  는  $i$  번째 채널의  $k$  번째 주파수 인덱스의 응답이다.  $k_1$  및  $k_2$  는 상호상관도 계산에서 고려하는 대역의 가장 낮은 주파수와 가장 높은 주파수를 나타낸다. 식 (1)의 정규화된 상호상관도는 시간 영역에서도 구할 수 있으나 본 연구에서는 주파수 영역에서의 빔포밍의 성능 향상을 목표로 하고 있으므로 간단한 디지털신호처리 기술을 도입하여 식 (1)과 같은 주파수 영역에서의 상관도를 추정한다.

기존의 전력비를 이용한 목표 신호 검출방법은 FBF 출력과 BM 출력 사이의 전력비를 이용해 목표 신호의 존재 유무를 결정하며 식 (2)과 같이 추정한다[7].

$$\phi_t^{(PR)} = \frac{\sum_{k=k_1}^{k_2} |Y_{FBF}(k)|^2}{\sum_{k=k_1}^{k_2} \sum_{m=2}^M |B_m(k)|^2} \quad (2)$$

여기서,  $B_m(k)$  는  $m$  번째 BM 출력의  $k$  번째 주파수 응답이다.  $M$ 은 입력 채널의 수를 나타낸다. 식 (1), (2) 에서 나타난  $\phi_t^{(NCC)}$ ,  $\phi_t^{(PR)}$  은 목표 음성 구간에서 큰 값을, 잡음 구간에서 작은 값을 나타내게 된다.

### 2.3. Gaussian Mixture Model (GMM)

GMM은 복잡한 형태를 갖는 함수를 가우시안 함수의 합으로 근사화시키는 데에 사용된다. GMM을 이용한 확률 분포는 식 (3)과 같이 M개의 가우시안 함수와 가중치로 정의된다[8].

$$P(\vec{x}|\lambda) = \sum_{i=1}^M w_i \mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i) \quad (3)$$

여기서,  $\vec{x}$  는 특징 벡터,  $w_i$  는  $i$  번째 가우시안 확률 분포의 혼합 가중치,  $M$ 은 가우시안 혼합의 개수,  $\mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i)$  는 D차원의 가우시안 분포로서 식 (4)와 같이 정의한다. 혼합 가중치는  $\sum_{i=1}^M w_i = 1$  을 만족한다.

$$\mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\} \quad (4)$$

여기서,  $\vec{\mu}_i$  는  $i$  번째 가우시안 분포의 D차 평균벡터,  $\Sigma_i$  는  $i$  번째 가우시안 분포의  $D \times D$  크기의 공분산행렬이다.  $\lambda$  는 각각의 가우시안 분포의 가중치, 평균벡터, 공분산행렬로 구성된 모델 파라미터로 식 (5)과 같이 나타낸다.

$$\lambda = \{w_i, \vec{\mu}_i, \Sigma_i\} \quad i = 1, \dots, M \quad (5)$$

각 모델 파라미터를 추정하는 훈련 DB에 expectation maximization (EM) 알고리즘을 적용하여 이루어진다 [8].

### III. 제안된 AMC 알고리즘

그림 2에서 제안된 AMC 알고리즘의 동작을 나타내는 블록 다이어그램을 나타내었다. 여기서  $B_i(k)$ 는 그림 1에서 나타낸 BM의  $i$  번째 출력 주파수 응답을 나타낸다. 주파수 밴드 단위에서 BM의 출력 에너지와 FBF의 출력 에너지를 구한 후, 그것의 전력비를 구하여 특징 벡터를 구성한다. 추출된 특징벡터를 이미 추정해 놓은 목표 음성구간 및 잡음 구간의 GMM과 비교하여 log-likelihood ratio (LR)을 추정하며,  $LR_{TH}$ 로 주어지는 임계치와 비교하여 목표 음성구간인지 순수 잡음구간인지를 판정한다. 중요 블록 다이어그램의 동작에 대한 설명은 다음과 같다.

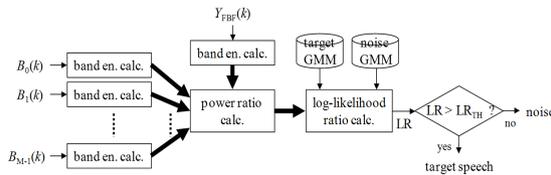


그림 2. 제안된 AMC 알고리즘의 블록 다이어그램  
Fig. 2 Block diagram of proposed algorithm

#### 3.1. 주파수 밴드 단위의 에너지 계산

식 (6)에서 주파수 밴드 단위의 에너지 계산 방법을 나타내었다.

$$P_{BMm}^{(l)} = \sum_{k=k_s^{(l)}}^{k_e^{(l)}} |B_m(k)|^2, \quad l = 0, 1, \dots, N_B - 1 \quad (6)$$

여기서,  $l$ ,  $k_s^{(l)}$ ,  $k_e^{(l)}$  은 각각 주파수 밴드 인덱스,  $l$  번째 밴드의 시작 주파수 인덱스,  $l$  번째 밴드의 끝 주파수 인덱스를 나타낸다.  $N_B$ 는 주파수 밴드의 개수이다. FBF 출력의 밴드별 에너지  $P_{FBF}^{(l)}$ 을 수식 (6)을 활용하여 구할 수 있다.

#### 3.2. 전력비의 계산

전력비는 주파수 밴드별 BM 출력 에너지의 채널별 합과 FBF 출력의 에너지 비를 구하며 식 (7)와 같이 계산한다.

$$PR^{(l)} = \frac{P_{FBF}^{(l)}}{\frac{1}{M-1} \sum_{m=2}^M P_{BMm}^{(l)}} \quad (7)$$

#### 3.3. 가우시안 혼합 모델을 이용한 스코어 계산

훈련 과정을 통해서 미리 구축된 목표 음성구간과 순수 잡음구간에서의 주파수 밴드 단위 전력비의 GMM 기반 확률 분포를 이용하여 식 (8)의 방식으로 LR을 측정한다.

$$LR = \log P(\vec{PR} | \lambda_T) - \log P(\vec{PR} | \lambda_N) \quad (8)$$

여기서,  $\vec{PR} = [PR^{(0)}, PR^{(1)}, \dots, PR^{(N_B-1)}]^T$ ,  $\lambda_T$ ,  $\lambda_N$ 는 목표 음성구간과 잡음 구간에서 훈련 과정에서 추정해 놓은 GMM 파라미터이다.

#### 3.4. 목표 음성의 검출

FBF의 출력은 목표 음성 구간일 때 큰 값을 갖고 순수 잡음구간일 때는 상대적으로 작은 값을 갖는다. 또한, BM의 출력은 목표 음성구간일 때 낮은 값을 갖고 순수 잡음구간일 때 상대적으로 작은 값을 갖는다. 따라서, 식 (8)의 LR은 목표 음성구간에서 높은 값을 갖게 되고, 잡음 구간에서 낮은 값을 갖게 될 것이다. 그림 2에서의  $LR_{TH}$ 는 훈련과정을 통해서 가장 낮은 검출 오류를 갖는 값으로 설정된다. 목표 음성구간을 검출하기 위한 결정 규칙을 식 (9)에 나타내었다.

$$LR > LR_{TH} \quad (9)$$

### IV. 실험 및 결과

#### 4.1. 다채널 DB 수집 및 실험 조건

마이크로폰은 8 개를 사용하여 그림 3과 같이 반지름이 8cm인 원형으로 배치하였고 마이크로폰 1번과 8번을 기준으로 정면 방향을 설정하였다. 목표 음성신호원은 정면에 위치하도록 하여 DB를 수집하였다. 모든 데이터는 표본화율 16 kHz, 해상도 16 bit로 8 채널 녹음 장치를 이용해 수집하였다. DB 수집을 위해서 9 m × 11 m × 3 m 크기의 대학교 중형 강의실을 활용하였다. 목

표 음성신호와 잡음 신호는 별도로 수집되어 원하는 SNR이 되도록 인공적으로 가산하여 활용하였다. 목표 음성신호로 발생당 지속시간이 약 1초인 총 1808개의 한국어 고립어를 고품질 스피커를 통하여 재생시켜서 수집하였다. 마이크론과 목표 음성신호간의 거리는 1.8 m이며 구체적인 수집 조건을 그림 4에 나타내었다. 다채널 잡음 데이터는 식당에서 수집한 배틀성 잡음을 6개의 스피커에서 재생하여 활용하였다. 그림 5에 수집 조건을 나타내었다. 수집된 잡음은 입력 SNR이 -5 dB, 0 dB, 5 dB가 되도록 목표 음성신호에 인공적으로 가산되었다. 최종적인 잡음환경의 DB 개수는  $3 \times 1808 = 5424$  이다. 그 중 50%는 훈련 과정에, 나머지 50%는 테스트 과정에 활용되었다.

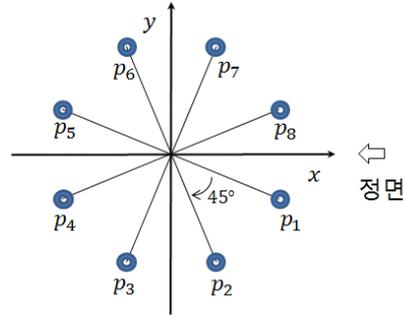


그림 3. 다채널 마이크론의 배치  
Fig. 3 Placement of multichannel microphones

#### 4.2. GSC 알고리즘

그림 1에서 나타낸 바와 같이 GSC 알고리즘은 DFT를 활용한 주파수 영역에서 구현되었다. 단구간 신호 분석을 위한 창함수의 크기 및 DFT의 크기는 모두 512였으며, 매 5 ms 단위로 주파수 분석을 수행하였다. BM의 출력을 구하기 위한 기준 채널로 1번을 활용하였다. NC 블록의 적응 필터 계수는 normalized LMS (NLMS) 알고리즘에 의해서 갱신되었다[9].

#### 4.3. 제안된 AMC 알고리즘

그림 2의 제안된 방식에 의한 AMC를 구현하기 위해서 1 kHz 단위의 주파수 밴드 분할을 수행하였다. 즉, 1개의 주파수 밴드당 총 32개의 DFT 주파수 성분을 고려하였다. 이를 바탕으로 식 (7), (8)의 8차의  $\overline{PR}$ 을 구성하였다. 잡음이 섞인 다채널 DB에서 잡음구간 및 목표 음성구간은 수작업에 의해서 표기되었다. 식 (8)의 계산을 위한 잡음구간 및 목표 음성구간의 확률 분포는 혼합의 수가 4인 GMM으로 모델링 되었다.

#### 4.4. 성능 평가

제안한 목표 신호 검출 알고리즘의 성능은 기존의 정규화된 상호상관도를 이용한 방법, 기존의 전력비를 이용한 방법과 비교되었다. 식 (1)의 정규화된 상호상관도를 계산하기 위해서 그림 3의 채널 2번과 채널 7번에서 입력된 신호가 사용되었다. 식 (1)과 (2)를 계산할 때 고려되었던 주파수 대역은 일반적으로 음성의 에너지가 가장 높게 분포한다고 알려진 250 Hz - 1 kHz 였다.

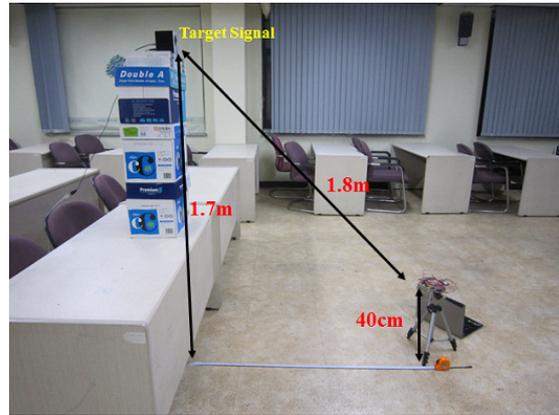
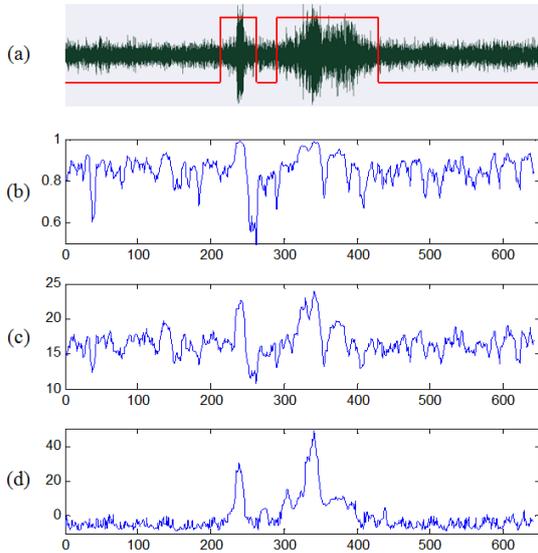


그림 4. 목표 신호 수집 환경  
Fig. 4 Environment of target signal collection



그림 5. 잡음 수집 환경  
Fig. 5 Environment of noise collection

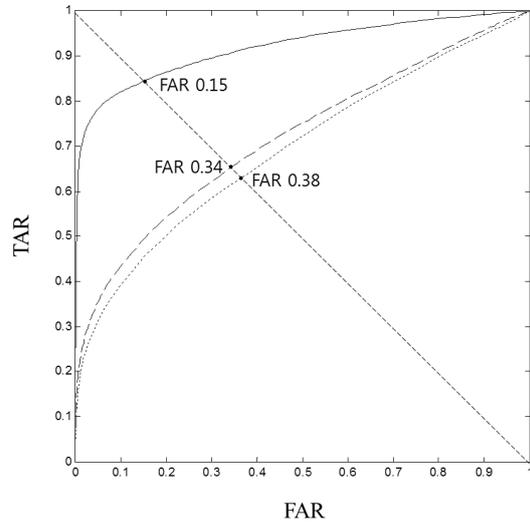


**그림 6.** 목표 신호 검출 기법들의 검출 궤적 비교(발성음: /소프트웨어/, 수평축: 프레임 단위의 시간) (a) 입력 SNR 5dB의 파형 및 목표 음성 구간(높은 레벨 구간이 목표 음성 구간), (b) 정규화된 상호 상관도, (c) 기존의 전력비, (d) 제안된 밴드 단위 전력비의 GMM 기반 LR

**Fig. 6** Comparison of detection contour of target signal detection methods(Utterance: "software", horizontal axis: time in frame unit) (a) Waveforms and target voice section of input SNR 5dB(High level is target voice section) (b) Normalized cross correlation, (C) Conventional power ratio, (d) LR based on GMM of proposed bandwidth power ratio

성능 지수로서 목표 음성구간과 순수 잡음구간을 얼마나 정확히 검출하느냐를 나타내는 receiver operating characteristic (ROC) 곡선과 각 알고리즘이 GSC의 AMC로 채택되었을 때의 출력 SNR이 측정되었다[10].

그림 7에 모든 훈련 DB를 활용하여 측정한 각 알고리즘별 ROC 곡선을 나타내었다. 그림 7의 FAR은 false acceptance rate로서 잡음구간을 목표 음성구간으로 분류할 확률, TAR은 true acceptance rate로서 목표 음성구간을 제대로 분류할 확률을 나타낸다. 그림 7에서 equal error rate (EER)에서의 FAR 및 임계치는 정규화된 상호상관도일 때 0.38, 0.87, 기존의 전력비를 사용할 때 0.34, 16.7, 제안된 방식을 사용할 때 0.15, -1.7로 측정되었다. 그림 7에서 알 수 있듯이 제안된 방식에 의한 목표 신호의 검출 성능이 종래의 알고리즘들에 비해서 우수하였다.



**그림 7.** 훈련 DB에서 측정한 ROC 곡선(점선: 정규화된 상호 상관도, 파선: 기존의 전력비, 실선: 제안된 방식)  
**Fig. 7** ROC curves measured from train DB(dotted: Normalized cross correlation, dashed: conventional power ratio, solid: proposed method)

이러한 성능이 측정되는 이유는 다음과 같이 분석이 가능하다. 식 (1)의 NCC의 경우 가용 채널이 많음에도 한 쌍의 채널만 활용하고 있다는 것, 대부분의 음향 신호의 분포가 저주파에 집중되어 있기 때문에 상호상관도의 값이 시간차를 갖는 순수 잡음 구간에서도 예상한 만큼 낮은 값을 갖지 않는다는 것 등의 이유로 성능이 낮다고 볼 수 있다. 식 (2)의 기존의 전력비 방식의 경우 모든 채널의 정보를 활용하기 때문에 NCC 보다는 검출 궤적이 좋은 특성을 갖지만 만족할 만한 성능에는 이르지 못한다고 볼 수 있다.

즉, 입력된 신호에 따라서 목표 신호의 검출에 가장 유리한 주파수 대역은 달라진다고 볼 수 있는데, 제안된 방식에서는 1 kHz 단위의 주파수 밴드에서 각각 특징을 추출하여 GMM에 의한 확률 모델링을 사용하기 때문에 기존의 고정된 주파수 대역만을 활용하는 방식에 비해서 더 좋은 성능을 나타내는 것으로 판단된다. 제안된 알고리즘에 대해서 훈련 DB의 EER에서 측정한 임계치를 테스트 DB에 적용하였을 때의 FAR 및 TAR을 표 1에 나타내었다.

그림 8에서 테스트 DB에 대한 GSC 출력의 SNR을 측정한 결과를 나타내었다.

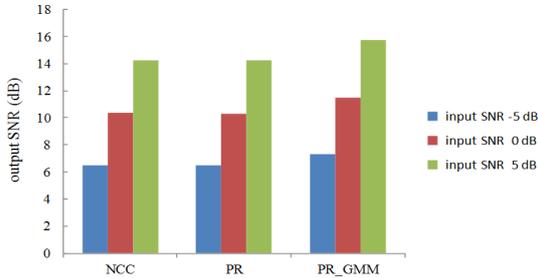


그림 8. 테스트 DB에 대해서 각 알고리즘을 GSC의 AMC로 활용하였을 때의 출력 SNR 측정 결과(NCC: 정규화된 상호상관도, PR: 기존의 전력비, PR\_GMM: 제안된 밴드 단위의 전력비의 GMM 기반 LR)

Fig. 8 Output SNR measurement results when utilizing each algorithm to AMC of GSC about test DB(NCC: Normalized Cross-correlation, PR: existing power ratio, PR\_GMM: LR based on GMM of proposed band unit power ratio)

표 1. 제안된 알고리즘의 테스트 DB에 대한 FAR과 TAR

Table. 1 FAR and TAR of the proposed algorithm for test DB

입력 SNR (dB)	FAR	TAR
-5	0.15	0.74
0	0.14	0.84
5	0.17	0.92

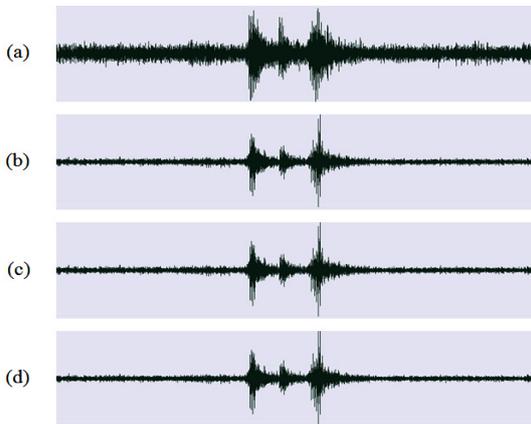


그림 9. GSC 빔포밍 전후의 알고리즘별 파형(발성음: /예컨대/)  
 (a) SNR 5 dB의 입력신호, (b) 정규화된 상호상관도 기반의 AMC, (c) 기존의 전력비 기반의 AMC, (d) 제안된 방식의 AMC  
 Fig. 9 Waveform of each algorithm before and after GSC beamforming(vocalization sound: /YeKeondae/) (a) Input signal of SNR 5 dB, (b) AMC based on Normalized Cross-correlation, (c) AMC based on existing power ratio, (d) Proposed AMC

출력 SNR의 측정은 목표 음성구간의 평균 에너지를 잡음 구간의 평균 에너지로 나눔에 의해서 측정되었다. 각각의 알고리즘에 의한 AMC 동작을 위하여 그림 7의 EER에서 추정된 임계치를 활용하였다. 그림 8에서 알 수 있듯이 입력 SNR에 상관없이 제안된 방식이 가장 좋은 성능을 보임을 알 수 있었다. 입력 SNR 5 dB에서 출력 SNR은 정규화된 상호상관도를 활용한 AMC 일 때 14.27 dB, 기존의 전력비일 때 14.28 dB, 제안된 방식일 때 15.7 dB 였다. 그림 9에서 입력 SNR이 5 dB일 때, 각 알고리즘이 GSC 빔포밍의 AMC로 활용되었을 때의 잡음제거 전후의 파형을 나타내었다.

## V. 결론

본 논문에서는 GSC 기반 빔포밍의 AMC를 위한 주파수 밴드별 전력비 분포의 혼합 가우시안 모델을 이용한 목표 신호 검출 기법을 제안하였다. 실험결과 제안한 방법에 의한 목표 신호의 검출 성능은 EER 관점에서 0.15로 가장 낮았고, GSC 빔포밍에 의한 잡음 제거 후의 SNR 이득이 가장 높았다. 향후, 제안된 알고리즘과 추가적인 특징 파라미터를 활용하여 주파수 밴드 단위로 동작하는 AMC 알고리즘을 구현할 예정이다.

## 감사의 글

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

## REFERENCES

- [ 1 ] ETSI ES 202 212, Speech processing, transmission and quality aspects (STQ), v.1.1.2, 2005.
- [ 2 ] S. Jeong and M. Hahn, "Speech quality and recognition rate improvement in car noise environments," *Electronics Letters*, Vol.37, No.12, pp. 801-802, 2001.
- [ 3 ] A. Hyvarinen and E. Oja, "Independent component analysis:

- Algorithms and applications,” *Neural Networks*, vol. 13, no. 4, pp. 411-430, 2000.
- [ 4 ] O. Frost, “An algorithm for linearly constrained adaptive array processing,” *Proceedings of the IEEE*, Vol 60, No. 8, pp. 926-935, 1972.
- [ 5 ] S. Gannot et al., “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*, Vol. 49, No. 8, pp. 1614-1626, 2001.
- [ 6 ] Y. Jung, H. Kang, C. Lee, D. Youn, C. Choi, and J. Kim, “Adaptive microphone array system with two-stage adaptation mode controller,” *IEICE Trans. Fund.*, vol. E88-A, no. 4, pp. 972-977, Apr. 2005.
- [ 7 ] O. Hoshuyama, A. Sugiyama, and A. Hirano, “A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” *IEEE Trans. Signal Process.*, Vol 47, No. 10, pp. 2677-2684, 1999.
- [ 8 ] L. Rabiner and B. Juang, *Fundamentals of Speech Recognitions*, Prentice Hall, 1993.
- [ 9 ] M. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, 1996.
- [10] F. Tom, “An introduction to ROC analysis”, *Pattern Recognition Letters*, Vol. 27, pp. 861-874. 2006.



**장형욱(Hyungwook Chang)**

2013년 경상대학교 전자공학과(학사)  
2013년 - 현재 경상대학교 전자공학과(석사과정)  
※관심분야 : 음성신호처리, real-time operating system



**김영일(Youngil Kim)**

1979년 경북대학교 전자공학과(학사)  
1981년 연세대학교 전자공학과(석사)  
1985년 연세대학교 전자공학과(박사)  
1987년 - 현재 경상대학교 전자공학과/공학연구원(교수)  
※관심분야 : 음향공학, 음성신호처리



**정상배(Sangbae Jeong)**

1997년 부산대학교 전자공학과(학사)  
1999년 한국과학기술원 전기 및 전자공학과(석사)  
2002년 한국과학기술원 정보통신공학과(박사)  
2002년 - 2006년 삼성종합기술원 컴퓨팅랩(책임연구원)  
2006년 - 2009년 한국과학기술원 디지털미디어랩(연구조교수)  
2009년 - 현재 경상대학교 전자공학과/공학연구원(부교수)  
※관심분야 : 음성신호처리, 음성오디오 부호화