빔포밍 및 DOA 기반의 마스킹을 이용한 2채널 잡음제거

김영일* · 정상배**

Two-Channel Noise Reduction Using Beamforming and DOA-Based Masking

Youngil Kim* · Sangbae Jeong**

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

요 약

본 논문에서는 빔포밍과 입사각 분석 기반 마스킹을 이용한 다채널 음성개선 알고리즘이 제안된다. 제안된 알고리즘에서는 LCMV 빔포밍을 수행한 후에 입사각 분석을 이용한 멜-주파수 위너필터가 적용되어 잔존하는 잡음을 제거한다. 성능 향상을 위해서 빔포밍의 적응 필터 학습률과 목표 음성 스펙트럼 검출을 위한 입사각 임계치가 최적화된다. 성능 지수로서 PESQ와 출력 SNR이 측정되었으며 실험 결과 제안한 알고리즘이 종전의 최소 분산 빔포밍 기법 보다 PESQ 관점에서 0.09, 출력 SNR 관점에서 5.75 dB의 성능 향상시킴을 알 수 있었다.

ABSTRACT

In this paper, we propose a multi-channel speech enhancement algorithm using beamforming and direction-of-arrival (DOA)-based masking. The proposed algorithm enhances noisy speech basically by the linearly constrained minimum variance (LCMV) algorithm and then a mel-scale Wiener filter designed using DOA-based masking is applied to remove still remaining noises. To improve the performance, we optimize the learning rate of the adaptive filters in LCMV and the DOA threshold to detect target speech spectrum. As performance indices, the perceptual evaluation of speech quality (PESQ) score and output SNRs are measured. Experimental results show that the proposed algorithm outperforms the conventional LCMV beamformer by 0.09 in PESQ score and 5.75 dB in output SNR, respectively.

키워드

잡음제거, 선형 제약 최소 분산, 입사각, 빔포밍

Key word

noise reduction, linearly constrained minimum variance, direction-of-arrival, beamforming

* 정회원 : 경상대학교
** 정회원 : 경상대학교 (교신저자, jeongsb@gnu.ac.kr)

접수일자 : 2012. 07. 30
심사완료일자 : 2012. 08. 17

I. 서 론

음성은 기계-인간 인터페이스 중에서 사용자들에게 큰 편리성을 제공할 수 있는 수단이기 때문에 지속적으로 연구되어왔다. 그러나 음성 인터페이스의 성능은 주변 환경 잡음에 많이 영향을 받는 문제점을 가진다. 잡음이 존재하면 통화품질 또는 음성인식기의 음성인식률이 크게 저하될 수 있다. 이러한 문제점을 해결하기 위하여, 음성 인터페이스의 성능 개선을 위한 다양한 음성 개선 기술들이 제안되어 왔다. 음성개선을 위한 잡음 제거기술은 잡음의 종류에 따라 달라진다. 일반적으로 정상성 잡음을 제거하는 방법으로는 단일채널 기반의 위너필터와 칼만필터 방법이 있다[1][2]. 반면에 비정상성 잡음을 제거하기 위한 방법으로는 다채널 기반 빔포밍(beamforming)과 맹목신호분리가 있다[3-6][7][8]. 일반적으로, 빔포밍 기술은 주로 로봇이나 TV에 음성 인터페이스를 적용시키기 위한 연구가 수행되어 왔으며 4-채널 이상을 사용하고 있다. 반면에 이동형 단말기를 위한 빔포밍 기법은 하드웨어 상의 제약조건 때문에 2-채널 이하를 사용하는 것이 불가피하다. 맹목신호분리는 실시간 연산량의 문제와 사용하는 마이크로폰의 개수가 분리하고자 하는 음원의 개수에 의존하는 문제 때문에 실제 적용에 있어서 신뢰할 만한 성능을 발휘하는데 어려움이 따른다[4][6].

이에 반해, 빔포밍은 목표 음성신호의 위치정보를 미리 알 수 있기 때문에 실제 적용에 있어서 맹목신호분리보다 더 좋은 성능을 발휘할 수 있다고 알려져 있다[9]. 가장 널리 사용되고 있는 빔포밍 알고리즘으로는 LCMV와 generalized sidelobe canceller (GSC)가 있다[3][5][7-9]. GSC는 목표 음성신호 구간에서는 잡음 제거를 수행하지 않기 때문에 마이크로폰의 개수를 2개 이하로 사용해야 하는 이동형 단말기에 적용시키기에 부적합한 측면이 있다. 하지만 LCMV 알고리즘의 경우에는 목표 음성신호 구간에서도 잡음 제거를 수행하기 때문에 마이크로폰의 개수가 2개 이하로 제약되어 있을 때 음질측면에서는 GSC보다 더 나은 성능을 기대할 수 있다.

본 논문에서 제안하는 음성개선 알고리즘의 목적은 휴대용 단말기에 적용시키는 것과 목표 음성신호 구간에서 음질을 개선하는 것이다. 이를 위하여, 2-채널 마이크로폰을 사용하였고 참고문헌[10]의 LCMV 알고리

즘을 기반으로 하였다. LCMV에서 학습률이 이상적인 학습률 보다 낮으면 목표신호에서 충분하게 잡음제거가 되지 않기 때문에 좋은 음질을 얻을 수 없다. 반면에 학습률이 이상적인 학습률 보다 높으면 목표신호에서 과도하게 잡음제거를 수행하기 때문에 잡음은 많이 제거되지만 목표 음성신호까지 왜곡시키게 되므로 결과적으로 낮은 학습률과 마찬가지로 좋은 음질을 얻을 수 없게 된다. 그러므로 LCMV과정에서는 음질측면에서 목표신호의 왜곡과 잡음 제거량이 고려된 최적의 학습률을 적용시켜야 한다. 이러한 최적의 학습률을 얻기 위하여 본 논문에서는 음질을 성능지수로 한 전수조사 방식을 수행하였다. 그리고 추가적인 잡음제거를 위하여 마스크 기법을 이용하여 추정된 위너필터를 LCMV의 후처리 방법으로 사용하였다[11][12]. 마스크는 일반적으로 채널간 강도차나 채널간 시간차를 이용함으로써 잡음인지 아닌지를 판별하여 목표 음성신호와 잡음을 이진 분리하여 잡음부분을 상쇄시키는 개념의 잡음제거 방법이다.

본 논문에서는 마스크를 위한 파라미터로 입사각(direction-of-arrival (DOA))를 사용하였으며 잡음과 목표신호를 분리한 후 멜-주파수 위너필터를 통하여 추가적으로 잡음을 제거하였다.

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

II. 관련 연구

빔포밍은 방향정보를 이용하여 인간의 음성 및 음악 신호와 같은 비정상성 잡음을 제거한다. 빔포밍 기술은 크게 LCMV와 GSC로 나뉘어 연구되어 왔다. 그리고 마스크는 입력신호의 특정 주파수 성분이 목표 음성인지 혹은 잡음인지를 판별하는 역할을 수행한다. 각 알고리즘에 대한 자세한 설명은 다음과 같다.

2.1. Generalized Sidelobe Canceller

GSC 알고리즘은 크게 고정빔포밍부, 차단행렬부, 잡음제거부의 블록으로 구성된다. 고정빔포밍부에서는 목표 음성신호 방향으로 신호의 이득과 시간차를 보정

하는 역할을 수행한다. 차단행렬부에서는 목표 음성신호원의 방향정보를 바탕으로 참조 잡음신호를 생성한다. 잡음제거부에서는 고정빔포밍부의 출력에 존재하는 잔존 잡음을 차단행렬부에서 추정한 참조잡음신호와 적응필터를 이용하여 제거하는 역할을 한다. 일반적으로 목표 음성신호와 비정상성 잡음의 경우 주파수의 분포가 유사하므로 음질을 보존하기 위하여 적응 모드 제어기를 사용한다[8].

2.2. Linearly Constrained Minimum Variance

LCMV의 기본 원리는 정면방향으로 선형조건을 두어 출력의 전력을 최소화 하는 것이다. 그림 1은 LCMV 빔포머의 기본적인 구조를 나타내고 있다[10]. 그림 1에서 K와 J는 각각 마이크로폰의 개수와 적응필터의 길이이다. LCMV 알고리즘에서는 목표 음원이 항상 마이크로폰 배열의 정면에 위치한다고 가정한다. 따라서, 정면이 아닌 다른 방향의 신호는 잡음으로 간주한다.

그림 1. LCMV 빔포머의 기본적인 구조
Fig. 1 The basic structure of LCMV beamformer

그림 1의 맨 아래 부분은 목표 음성신호원에 대한 LCMV 빔포머의 등가 필터 구조를 보여주고 있다. LCMV 빔포밍의 기본 방식은 빔포머의 출력을 최소화 하는 동시에 목표 음원의 방향에 대한 주파수 응답은

항상 1이 되도록 하는 것이다. 이러한 동작을 식 (1)에 나타내었다.

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

식 (1)에서 $\vec{w} = [w_1, w_2, \dots, w_K]$ 이고, $E[\cdot]$ 는 통계적 평균 연산자이다. $y(n)$ 과 $f(n)$ 은 각각 빔포머의 출력과 등가 필터 가중치이다. 식 (1)은 라그랑지 승수를 이용하여 최적화 된다. 최종적으로 식 (2)로 정리된다[10].

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

식 (2)에서 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. C는 조건 행렬이고 $\vec{F} = [1, 0, \dots, 0]$ 이다. 하지만 식 (2)를 실제로 구현하는데 있어서 자기상관 행렬의 역행렬을 구하기 위해서는 많은 연산량을 필요로 한다. 따라서, 실제 구현을 할 때는 식 (3)을 이용하여 구현한다[10].

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

식 (3)에서 μ 는 적응필터의 학습률, $P = I - C(C^T C)^{-1} C^T$ 이다.

2.3. DOA 기반 마스크

본 연구에서는 DOA 기반의 마스크를 구현하였었다. 그림 2에서 DOA의 개념도를 나타내었다[13]. 그림 2에서 d, L, θ 는 마이크로폰의 간격, 잡음이 기준채널과 다른 채널에 도달했을 때의 거리 그리고 DOA이다. θ 의 유도과정은 식 (4)에서 시작된다.

$$\Delta n = F_s \Delta t = \frac{d F_s}{v_s} \sin \theta \quad (4)$$

그림 2. DOA의 개념
Fig. 2 The concept of DOA

여기서 Δn , F_s , Δt , v_s 는 샘플단위의 시간지연, 표본화율, 초단위의 시간지연, 음속이다. 1번 채널 입력신호와 2번 채널 입력신호의 주파수 응답 비는 식 (5)와 같다.

$$R = \frac{X_1(k)}{X_2(k)} = e^{j\frac{2\pi}{N}k\Delta n} \quad (5)$$

식 (5)에서 N , k , $X_1(k)$, $X_2(k)$ 는 이산 푸리에 변환 (discrete Fourier transform (DFT))의 크기, 주파수 인덱스, $x_1(n)$ 의 DFT 결과, $x_2(n)$ 의 DFT 결과이다. R 의 위상은 식 (6)과 같다.

$$\Phi_k = \frac{2\pi}{N}k\Delta n = \tan^{-1}\left(\frac{\text{Im}[R]}{\text{Re}[R]}\right) \quad (6)$$

식 (6)에서 $\text{Im}[\cdot]$, $\text{Re}[\cdot]$ 는 복소수의 허수부 및 실수부를 추출하는 연산자이다. 식 (6)에 식 (4)을 대입하면 식 (7)을 얻게 된다.

$$\Phi_k(k) = \frac{2\pi}{N}k\Delta n = \frac{2\pi}{N}k\frac{dF_s}{v_s}\sin\theta_k \quad (7)$$

식 (7)에서 θ_k 는 DFT 주파수 k 에서의 DOA를 나타낸다. 양변을 θ_k 에 대해서 정리하면 최종적으로 식 (8)을 얻을 수 있다.

$$\theta_k = \sin^{-1}\left(\Phi_k \frac{Nv_s}{2\pi kdF_s}\right) \quad (8)$$

본 연구에서는 식 (8)의 DOA 정보를 모든 주파수에 대해서 측정 후, 해당 주파수 성분이 임계치 이하일 경우에 그것을 목표 음성신호의 성분이라고 판정하고 그렇지 않을 경우 잡음으로 판정하는 마스크 기법을 적용하여 빔포밍 후의 추가적인 잡음제거에 활용한다.

III. 제안된 알고리즘

제안된 알고리즘의 목표는 기존의 LCMV 빔포머에 마스크를 접목시켜 효과적으로 잡음 제거를 하는 것이다. 즉, LCMV를 하여 기본적인 잡음제거가 된 결과를 입력신호의 DOA 분석 정보를 이용하여 목표 음성신호와 잡음을 분리한 후 위너필터를 통하여 추가적으로 잡음제거를 수행하는 개념이다. 이러한 동작을 위해서 본 논문에서는 채널간 위상차를 이용하여 DOA를 측정한다. 그림 3은 제안된 알고리즘의 전체적인 구조를 나타내고 있다.

그림 3. 제안된 알고리즘의 구조
Fig. 3 The structure of the proposed algorithm

제 1장 서론에서 언급한 바와 같이 본 연구에서는 이동형 단말기를 위한 잡음제거 알고리즘을 구현하는 것이므로 마이크론의 개수는 2 개를 사용하는 것으로 한다. 그림 3에서 $x_1(n)$ 과 $x_2(n)$ 은 2-채널 입력 신호이다. $A(z)$ 는 채널보정 필터, $w_1(z)$, $w_2(z)$ 는 LCMV의 유한응답 적응필터, $x_{2,eq}(n)$ 는 $x_2(n)$ 의 채널보상 결과, $X_1(k)$ 과 $X_2(k)$ 는 $x_1(n)$ 과 $x_{2,eq}(n)$ 의 DFT 결과이다. θ_k 는 주파수 k 에서의 DOA, $y(n)$ 은 LCMV의 출력신호, $\hat{S}(k)$ 과 $\hat{V}(k)$ 는 마스크를 통하여 분리된 목표 음성신호와 잡음의 주파수 응답이다. 마지막으로 $y_{post}(n)$ 은 후처리에 의해서 잡음이 제거된 최종 목표 음성신호이다. 주요 블록에 대한 설명은 다음과 같다.

3.1. 채널 보정

2-채널의 마이크로폰으로 신호를 받아들일 경우, 마이크로폰 및 입력 장치의 특성 차이 때문에 생기는 채널 간의 주파수 응답 차이가 필연적으로 나타난다. 이때, LCMV는 정면 방향으로의 신호만 유지시키기 때문에 채널간 응답 차는 목표 음성신호를 왜곡시키는 결과를 초래할 수 있으므로 빔포밍의 성능이 저하될 수 있다. 따라서, 빔포밍의 성능 저하를 방지하기 위한 채널 보정이 필수적이며 이를 위한 비용함수를 식 (9)에 정의하였다[11].

$$J(k) = \sum_{t=0}^{T-1} |X_1(t, k) - A(k)X_2(t, k)|^2 \quad (9)$$

식 (9)에서 k , t , 그리고 T 는 각각 주파수 인덱스, 시간 인덱스, 목표 음성신호 프레임의 개수이다. $X_1(t, k)$ 과 $X_2(t, k)$ 는 정면의 목표 음성신호만이 존재할 때의 채널 1과 채널 2의 DFT 결과이다. $A(k)$ 는 채널 보상함수의 주파수 응답이다. 비용함수를 $A(k)$ 에 대해서 최적화 하면 식 (10)을 얻게 된다.

$$A^{(opt)}(k) = \frac{\sum_{t=0}^{T-1} X_1(t, k)X_2^*(t, k)}{\sum_{t=0}^{T-1} |X_1(t, k)|^2} \quad (10)$$

식 (10)에서 '*'는 복소공액 연산자이다. 최종적으로 채널 보상된 신호는 식 (11)에 의해서 구해진다.

$$x_{2,eq}(n) = \sum_{k=0}^{L-1} a^{(opt)}(k)x_2(n-k) \quad (11)$$

식 (11)에서 $a^{(opt)}(k)$ 는 식 (9)의 결과를 역 이산 푸리에 변환한 것, L 은 채널 보상 필터의 시간 응답의 길이이다.

3.2. 마스크 및 멜-주파수 위너필터

마스크에서는 주파수에 대한 DOA를 분석 정보를 이용하여 목표 음성신호와 잡음을 분리하게 된다. 목표 음성신호와 잡음을 분리하는 식을 (12)에 나타내었다.

$$\hat{S}(k) = \begin{cases} Y(k) & , \quad \theta_k \leq \theta_{TH} \\ 0 & , \quad otherwise \end{cases} \quad (12)$$

식 (12)에서 θ_{TH} 는 목표 음성신호와 잡음을 판별하기 위한 DOA의 임계치이다. 각 스펙트럼에서 할당되지 않은 주파수 성분의 값은 0으로 둔다. 마찬가지로, $\hat{V}(k)$ 도 식 (12)와 유사하게 추정한다. 즉, 임의의 주파수에서 DOA가 임계치 이상이면 $Y(k)$ 를 잡음으로 분류하며 그렇지 않으면 0을 할당한다. 마스크를 통하여 추정된 목표 음성신호와 잡음을 이용하여 멜-주파수 위너필터를 수행하게 된다[2]. 일반적인 위너필터는 주파수대역을 선형적으로 두고 필터 계수를 추정하지만 멜-주파수 위너필터는 사람의 귀의 인지적 특성이 반영된 것이기 때문에 동일한 필터 길이를 쓰더라도 잡음제거 후의 음질이 더욱 향상될 수 있다. 마스크에서 추정된 목표 음성신호와 잡음의 SNR을 이용하여 주파수 영역에서의 필터의 이득을 식 (13)과 같이 추정할 수 있다.

$$H_{mel} = \frac{SNR_{mel}(\tau)}{1 + SNR_{mel}(\tau)} \quad (13)$$

식 (13)에서 τ 와 $SNR_{mel}(\tau)$ 는 멜-주파수 필터뱅크의 인덱스와 τ 번째 멜-주파수 필터뱅크에서의 SNR이

다. 시간 영역의 위너필터 계수를 구하기 위해서 각 필터뱅크가 직사각형의 모양이라는 가정을 하여 식 (14), (15)를 이용한다[11].

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega \quad (14)$$

$$h(n) \cong \sum_{\tau=0}^{K-1} H_{mel}(\tau) \cos(\omega_{c,\tau} n) \frac{\Delta\tau}{F_s} \quad (15)$$

식 (14)에서 ω 는 $2\pi f$ 이다. 식 (15)에서 $\omega_{c,\tau}$ 와 $\Delta\tau$ 는 각각 $2\pi f_{c,\tau}/F_s$ 와 $f_{c,\tau+1} - f_{c,\tau-1}$ 과 같다. 여기서 $f_{c,\tau}$ 는 주파수 영역에서 τ 번째 멜-주파수 필터뱅크의 중심 주파수이다. 최종적으로 LCMV의 출력신호와 시간영역 위너필터의 합성곱을 수행함으로써 잡음이 제거된 목표 음성신호를 얻게 된다.

IV. 실험 및 결과

4.1. DB 구성

실험을 위해 사용된 모든 데이터베이스는 사무실 환경에서 스피커로 재생되는 신호를 고품질 디지털 녹음기를 사용하여 간격 2cm인 2채널 마이크로폰으로 수집되었다. 사용된 표본화율은 16 kHz였다. 목표 음성신호는 마이크로폰의 정면 30cm 거리에서 재생된 한국인 남녀 각 20명이 발성한 총 60문장이며 문장당 지속시간은 2~5초 정도였다. 잡음으로는 정면 기준 60도 방향 1.5m 거리에서 재생된 음악신호를 사용하였다. 그림 4는 DB 수집 환경을 나타내고 있다. 본 연구에서는 휴대용 단말기를 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 별도로 녹음되며 두 신호를 가산하여 원하는 SNR을 맞추었다. 본 논문에서 고려된 입력 신호의 SNR은 0 ~ 20 dB이다.

그림 4. DB 수집 환경
Fig. 4 Environment for DB collection

4.2. 성능지수

제안된 알고리즘의 성능평가를 위하여, PESQ와 출력 SNR을 성능지수로 하여 SNR별로 결과를 도출하였다. 여기서 PESQ는 ITU-T recommendation P. 862 (02/01)로서 표준화로 되어있다[14]. PESQ의 결과는 주로 1에서 5 사이의 값을 가지는 mean opinion score (MOS)로 모델링된다. 또 다른 성능지수인 SNR은 음성구간과 잡음구간의 전력비를 나눈 비를 이용하였다.

4.3. 실험 결과 및 분석

실험을 위하여 LCMV의 적응 필터의 길이를 128, 마스킹을 위한 고속 푸리에 변환 크기를 512, 멜-주파수 필터뱅크의 개수는 33개를 사용하였다. 최적화된 기존의 LCMV와 제안된 알고리즘과의 성능비교를 위하여 모든 입력 SNR에 대한 평균 PESQ를 성능지수로 하여 학습률과 DOA 임계치의 최적화 과정을 수행하였다. 첫 번째로, 기존 LCMV에서 적응필터의 학습률을 변화시켰을 때 잡음제거 후의 모든 DB에 대한 평균 PESQ를 측정하였으며 그 결과를 그림 5에 나타내었다. 적응필터의 학습률이 0.5일 때 가장 좋은 성능을 보였다. 두 번째로, 제안된 방식에서 LCMV 적응필터의 학습률을 0.5로 할당한 후에 마스킹의 DOA 임계치를 변경하였을 때의 평균 PESQ를 측정하여 그 결과를 그림 6에 나타내었다.

그림 5. 학습률에 따른 기존 LCMV의 성능
Fig. 5 Performance of the conventional LCMV beamformer according to its filter learning

그림 8. 입력 SNR에 따른 출력 SNR의 비교
Fig. 8 Comparative output SNRs according to the input SNRs

그림 6. DOA 임계치에 따른 제안된 알고리즘의 성능
Fig. 6 Performance of the proposed algorithm according to the DOA threshold

그림 7. 입력 SNR에 따른 PESQ 비교 (성능지수 : PESQ, '●' : 제안된 알고리즘, '◆' : 기존의 LCMV)
Fig. 7 Comparative PESQ scores according to the input SNRs ('●' : Proposed algorithm, '◆' : conventional LCMV)

그림 9. 잡음제거 후의 결과 파형; (a) 입력신호 (/내일 날씨정보 알려줘/, 입력 SNR 5dB) (b) 기존의 LCMV 결과(필터 학습률 0.5) (c) 제안된 LCMV 결과(필터 학습률 0.5, DOA 임계치 15도)

Fig. 9 Waveforms after noise reduction (a) Noisy input (/내일 날씨정보 알려줘/, input SNR 5 dB) (b) Conventional LCMV result (filter learning rate 0.5) (c) Proposed LCMV result (filter learning rate 0.5, DOA threshold 15 degree)

DOA 임계치가 15도 일 때 가장 좋은 성능을 보임을 알 수 있었다. 세 번째로, 필터 적응률과 DOA 임계치를 최적의 값으로 할당하였을 때 입력 SNR에 따른 PESQ와 출력 SNR을 측정하였으며 그 결과를 그림 7, 8에 나타내었다.

제안된 알고리즘이 기존의 LCMV 빔포밍에 비해서 모든 입력 SNR에 대해서 더 나은 성능을 나타내었으며 입력 SNR 5 dB에서 PESQ 기준 0.15, 출력 SNR 기준 5.39 dB 향상되었다. 추가적으로 그림 9에 기존의 LCMV 방식 및 제안된 방식을 사용했을 때의 잡음제거 후의 결과 파형을 나타내었다. 종합적인 분석을 통해서 제안된 알고리즘이 잡음 제거에 매우 효과적임을 재차 확인할 수 있었다. 제안된 알고리즘이 더 좋은 성능을 나타낸 이유는 후처리로 채택된 마스킹 기반의 멜-주파수 위너필터 때문으로 볼 수 있다.

즉, 목표 음원과 마이크로폰간의 거리가 짧을 때는 DOA 임계치에 의한 주파수 축에서의 목표 음성신호 성분 검출이 용이하며 이를 기반으로 한 잡음제거가 매우 효과적이었다.

V. 결 론

본 논문에서는 DOA 기반 마스킹을 이용한 2채널 음성개선 알고리즘을 제안하였다. 최적화된 기존의 LCMV와 DOA 기반 마스킹을 하기위하여 모든 입력 SNR 기준으로 학습률과 DOA 임계치는 PESQ를 성능지수로 한 전수조사 방식을 사용하여 최적화 되었다. 최적 조건에서의 각 입력 SNR에 대한 PESQ 및 출력 SNR을 분석하였을 때 제안된 방식이 더 좋은 성능을 발휘함을 알 수 있었다. 향후 연구로 제안된 알고리즘에 음성검출기와 음성인식기를 결합하여 고립어 수준의 음성인식률 측정할 예정이다.

참고문헌

- [1] 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.
- [2] ES 202 212 V1.1.2 "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*, 2005.
- [3] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering", *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4-24, 1998.
- [4] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, 2001.
- [5] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing (Springer Topics in Signal Processing)*, Springer, 2008.
- [6] A. Hyvarinen, and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411-430, 2000.
- [7] 이영재, 김수환, 한승호, 한민수, 김영일, 정상배, "확률적 목표 음성 검출을 통한 다채널 입력 기반 음성개선," *한국음성학회 학술지 말소리와 음성과학*, 1권, 3호, pp. 97-104, 2009.
- [8] 박지훈, 이성주, 홍정표, 정상배, 한민수 (2008). "필터뱅크 기반 프로스트 알고리즘을 이용한 빔포밍 최적화," *대한음성학회 학술지 말소리*, 66호, pp. 73-86, 2008.
- [9] L. Wang, H. Ding, and F. Yin, "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, 2010.
- [10] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926-935, 1972.
- [11] S. Jeong, H. Yang, and M. Hahn, "Two-channel noise reduction for robust speech recognition in car environments," *Electronics Letters*, vol. 44, no. 17, pp. 1042-1043, 2008.
- [12] S. Jeong, S. Lee, and M. Hahn, "Dual microphone-based speech enhancement by spectral classification and Wiener filtering," *Electronics Letters*, vol. 44, no. 3, pp. 253-254, 2008.
- [13] 김수환, 이영재, 김영일, 정상배, "DOA 기반 학습률 조절을 이용한 다채널 음성개선 알고리즘," *한국음성학회 학술지 말소리와 음성과학*, 3권, 3호, pp. 91-98, 2011.
- [14] <http://en.wikipedia.org/wiki/PESQ>

저자소개

김영일(Youngil Kim)

1979년 2월 경북대학교
전자공학과 졸업(공학사)
1981년 2월 연세대학교
전자공학과 졸업(공학석사)

1985년 2월 연세대학교 전자공학과 졸업(공학박사)
1987년 3월 ~ 현재 경상대학교 전자공학과/
공학연구원(교수)

※ 관심분야: 음향공학, 음성신호처리

정상배(Sangbae Jeong)

1997년 3월 부산대학교 전자공학과
졸업(공학사)
1999년 2월 한국과학기술원 전기 및
전자공학과 졸업(공학석사)

2002년 8월 한국정보통신대학교 공학부 졸업
(공학박사)

2002년 9월 ~ 2006년 2월 삼성종합기술원 책임연구원
2006년 3월 ~ 2009년 2월 한국정보통신대학교 디지털
미디어연구소(연구조교수)

2009년 3월 ~ 현재 경상대학교 전자공학과/
공학연구원(조교수)

※ 관심분야: 음성인식, 음성/오디오부호화