한국음향학회지 제24권 제7호 pp. 423-430 (2005)

차량환경에서 음성인식 성능 향상을 위한 마이크로폰 어레이 빔형성 기법

A Microphone Array Beamformer for the Performance Enhancement of Speech Recognizer in Car

한 철 희*, 강 홍 구*, 황 영 수**, 윤 대 희*
(Chul-Hee Han*, Hong-Goo Kang*, Youngsoo Hwang**, and Dae-Hee Youn*)

*연세대학교 전기전자공학과, **관동대학교 전자정보통신기술공학부

(접수일자: 2005년 8월 16일; 채택일자: 2005년 9월 20일)

본 논문에서는 차량환경에서 잔향과 근접장 효과에 의해 발생하는 목적 음성 신호의 왜곡을 감소시킬 수 있는 마이크로폰 어레이 빔형성 기법을 제안하였다. 온라인으로 추정하기 어려운 소스와 마이크간의 전달함수 대신 상대적으로 추정이 용이한 기준 마이크와 다른 마이크간의 상대전달함수를 조향 벡터로 이용함으로써, 원격장 모델의 조향 벡터를 이용한 빔형성기에 비해 목적 음성 신호의 왜곡을 감소시킬 수 있는 준최적 빔형성 기법을 제안하였다. 제안된 방법의 성능을 검증하기 위해, 실제 차량에서 녹음된 음성 DB를 구축하고, 이를 이용하여 HTK를 통한 음성인식 실험을 수행하였다. 음성인식 실험 결과 원격장 모델을 이용한 방법보다 인식률이 최대 15%까지 향상됨을 확인하였다.

핵심용어: 마이크로폰 어레이, 상대전달함수, Near-Field, MVDR 빔형성기, 음성 개선, 음성 인식, RTF-MVDR
투고분야: 음성처리 분야 (2.3)

In this paper, a microphone array beamforming algorithm that reduces the signal distortion caused by reverberation and near-field effect in car environment is proposed. When reverberation or near-field effect is present, an optimum beamformer should be constructed with a steering vector consisting of transfer functions between source and microphones, but it is generally difficult to estimate transfer functions on-line without knowledge of the source signal. Instead, a sub-optimal beamforming algorithm that reduces signal distortion is proposed. It is constructed with steering vectors consisting of relative transfer functions between reference sensor and other sensors. In order to evaluate the performance of the proposed algorithm, we had recorded noisy speech database in a car, and performed speech recognition experiments with HMM Toolkit (HTK) released by Cambridge University. The recognition rate of the proposed algorithm was 15 percents higher than that of the conventional far-field beamformers in best case.

Keywords: Microphone Array, Relative Transfer Function, Near-Field, MVDR Beamformer, Speech Enhancement, Speech Recognition, RTF-MVDR
ASK subject classification: Speech Signal Processing (2.3)

I. 서 론

다채널 빔형성 알고리듬은 시간적 정보 이외에 공간적 정보를 같이 이용하므로 지향성이 있는 잡음에 대해 우수한 제거 능력을 갖으며, 비정적인 잡음까지도 제거가 가능하다는 장점 때문에 널리 쓰인다[1]. 차량 환경에서 음성 인식 시스템을 위한 마이크로폰 어레이 음성 품질 개선에 많이 쓰이는 주파수 영역 MVDR (Minimum Variance Distortionless Response) 빔형성기는 신호를 주파수 영역으로 분해하여 각 주파수 bin을 협대역으로 가정한 후 협대역 MVDR 알고리듬을 적용한다[2-3]. 차량 환경에서는 잔향 (reverberation)과 근접장 (near-field) 효과 때문에 원격장 (far-field) 모델의 조향벡터 (steering vector)를 사용할 경우 목적 신호의 감쇄 및 왜곡이 생길 수 있다[4]. 이러한 왜곡을 줄이기 위해서는

책임저자: 한 철 희 (hch@dsp.yonsei.ac.kr)
120-749, 서울 서대문구 신촌동 134,
연세대학교 전기전자공학부 디지털 신호처리 연구실
(전화: 02-2123-4534; 팩스: 02-312-4584)

각도뿐만 아니라 거리정보까지 포함된 근접장 조향 벡터를 이용하거나, 미리 측정된 전달 함수를 조향 벡터로 이용하는 방법이 있다[4,5]. 전자는 잔향이 존재하는 경우나 어레이 지오메트리 오차가 존재하는 경우 성능이 저하되고, 후자는 룸 지오메트리가 바뀔때마다 미리 전달함수를 측정해야 하는 불편함이 있으며, 두 방법 모두 각도뿐만 아니라 거리까지 정확히 알아야 되기 때문에 시스템이 복잡해진다.

Affes 등은 화자의 위치가 좁은 영역에서 변할 때, 화자의 위치 변화에 따른 기대 신호원과 마이크 사이의 전달함수를 추적 (tracking)하고 이를 GSC (Generalized Sidelobe Canceller)에 적용한 신호 부공간 추적 알고리듬을 제안하였다[6]. 이 방법은 화자의 위치 변화에 따라 변하는 전달함수를 추적할 수 있는 장점이 있으나, 사전에 비교적 정확한 초기 전달함수 추정치를 알고 있어야 한다는 단점이 있다.

Gannot 등은 잔향이 존재하는 경우의 주파수 영역 최적 Frost 빔형성기와 그것의 GSC 형태를 유도한 후, 상대전달함수 (Relative Transfer Function)를 차단 행렬 (Blocking Matrix)과 고정 빔형성기 (Fixed Beamformer)에 적용한 TF-GSC를 제안하였다[3,7,8]. 이 방법은 출력단의 신호성분에 소스와 기준센서간의 전달함수만큼 왜곡이 생기므로 준최적 솔루션이기는 하지만, 사전 정보없이 상대전달함수를 온라인으로 추정할 수 있고 어레이 지오메트리에 오차가 있어도 성능에 영향이 적다는 장점을 가지고 있다.

본 논문에서는 차량환경에서 구현이 용이하고, 안정적이며 잔향 및 근접장 효과에 강인한 빔형성 알고리듬을 제안하였다. TF-GSC의 제한조건 (constraint) 으로부터 같은 제한조건을 갖는 주파수 영역 MVDR 빔형성기를 유도하였으며, 이것이 MVDR 빔형성기의 조향벡터로 전달함수 대신 상대전달함수를 갖는 것과 같음을 보였다. 또한, 유사한 방법을 적용하여 상대전달함수를 조향벡터로 갖는 Delay-and-Sum 빔형성기를 제안하였다. 제안한 알고리듬의 성능 비교를 위하여 자동차에서 녹음된 음성 데이터베이스와 HTK 라이브러리를 이용하여 음성인식 실험을 수행하였다[9]. 실험 결과, 상대전달함수를 조향벡터로 갖는 제안한 알고리듬이 원격장 조향벡터를 갖는 기존 알고리듬에 비해 우수한 성능을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 주파수 영역 MVDR 빔형성기와 TF-GSC 알고리듬에 대해 설명하고, 3장에서는 제안된 상대전달함수를 이용한 주파수 영역 빔형성 기법에 대한 설명을 하였고, 4장에서는 자동차에서의 음성 데이터베이스 녹음과정 및 실험결과를 설명하였고, 5장에서 결론을 맺는다.

II. 잔향이 존재하는 환경에서의 주파수 영역 빔형성 기법

2.1. 주파수 영역 MVDR 빔형성기

잔향 (reverberation)과 잡음 및 간섭신호가 존재하는 환경에서 L개의 센서를 갖는 등간격 선형 마이크로폰 어레이 (uniform linear microphone array)가 있다고 할 때, 다음과 같은 신호 모델을 가정해보자.

$$x_l(t) = a_l(t) * s(t) + n_l(t), \quad l = 1, \ \dots \ , L \tag{1}$$

여기서, $x_l(t)$는 l번째 센서입력 신호, $s(t)$와 $n_l(t)$는 각각 목적 신호와 l번째 센서에 유입되는 잡음 및 간섭신호를 나타내며, $a_l(t)$는 목적 신호원과 l번째 센서의 전달함수를 나타낸다. 식 (1)을 단구간 퓨리에 변환 (STFT; Short Time Fourier Transform)하면 분석 창 (analysis window)의 길이가 충분히 클 때 식 (2)를 근사적으로 만족한다[3].

$$X_l(t,e^{j\omega}) \approx A_l(t,e^{j\omega})\, S(t,e^{j\omega}) + N_l(t,e^{j\omega}), \quad l = 1, \ \dots \ , L \tag{2}$$

식 (2)를 벡터 형태로 나타내면 식 (3)과 같다.

$$\mathbf{X}(t,e^{j\omega}) = \mathbf{A}(t,e^{j\omega})\, S(t,e^{j\omega}) + \mathbf{N}(t,e^{j\omega}) \tag{3}$$

여기서,

$$\mathbf{X}(t,e^{j\omega}) = \left[X_1(t,e^{j\omega}) \ X_2(t,e^{j\omega}) \ \cdots \ X_L(t,e^{j\omega}) \right]^T$$

$$\mathbf{A}(t,e^{j\omega}) = \left[A_1(t,e^{j\omega}) \ A_2(t,e^{j\omega}) \ \cdots \ A_L(t,e^{j\omega}) \right]^T$$

$$\mathbf{N}(t,e^{j\omega}) = \left[N_1(t,e^{j\omega}) \ N_2(t,e^{j\omega}) \ \cdots \ N_L(t,e^{j\omega}) \right]^T .$$

식 (3)은 협대역 주파수 스냅샷 모델과 같으므로 협대역에서와 같은 방법으로 최적 MVDR 빔형성기를 유도할 수 있다[2,3].

최적 MVDR 빔형성기의 출력신호는 식 (4)로 나타낼 수 있다.

$$Y(t,e^{j\omega}) = \mathbf{W}^H(t,e^{j\omega})\mathbf{X}(t,e^{j\omega}) \quad (4)$$

여기서,

$$\mathbf{W}(t,e^{j\omega}) = \left[W_1(t,e^{j\omega})\, W_2(t,e^{j\omega}) \cdots W_L(t,e^{j\omega})\right]^T.$$

무왜곡 제한조건 (distortionless constraint)을 만족하는 MVDR 빔형성기의 계수벡터는 식 (5)를 풀면 구할 수 있다.

$$\min_{\mathbf{W}}\left\{\mathbf{W}^H(t,e^{j\omega})\mathbf{\Phi}_{XX}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right\}$$
$$\text{subject to } \mathbf{W}^H(t,e^{j\omega})\mathbf{A}(t,e^{j\omega}) = 1. \quad (5)$$

여기서, $\mathbf{\Phi}_{XX}(t,e^{j\omega})$는 잡음구간의 공간상관행렬이고, $\mathbf{A}(t,e^{j\omega})$는 전달함수들로 이루어진 조향 벡터이다.

식 (5)을 만족하는 최적 MVDR 빔형성기의 계수는, 식 (6)으로 나타내어 진다[2,3].

$$\mathbf{W}_{opt,MVDR}(t,e^{j\omega}) = \frac{\mathbf{\Phi}_{XX}^{-1}(t,e^{j\omega})\mathbf{A}(t,e^{j\omega})}{\mathbf{A}^H(t,e^{j\omega})\mathbf{\Phi}_{XX}^{-1}(t,e^{j\omega})\mathbf{A}(t,e^{j\omega})} \quad (6)$$

2.2. TF-GSC 알고리듬[3]

신호원과 마이크간의 전달함수를 알 수 있으며, 2.1절에서와 같이 이것을 조향 벡터로 사용한다면 최적 MVDR 빔형성기를 구성할 수 있다. 그러나, 신호원과 마이크간의 전달함수를 구하려면 일반적으로 신호원을 알고 있어야 하는데 이는 온라인으로 추정하기 불가능하므로 오프라인에서 미리 측정해 놓아야 하며, 또한 시스템의 위치가 변하거나 사용공간이 변하면 매번 오프라인에서 전달함수를 다시 측정해야 한다는 제약이 있다.

Gannot 등은 식 (7)과 같은 상대전달함수를 정의하고 이를 고정 빔형성기 (Fixed Beamformer) 및 차단 행렬 (Blocking Matrix)에 사용하는 TF-GSC 알고리듬을 유도하였다.

$$H_l(t,e^{j\omega}) = \frac{A_l(t,e^{j\omega})}{A_r(t,e^{j\omega})}, \quad l = 1, \cdots, L \ (1 \le r \le L). \quad (7)$$

$H_l(t,e^{j\omega})$은 r번째 센서로부터 l번째 센서까지의 전달함수를 의미하며, 신호원과 센서간의 전달함수와 구별하기 위하여 상대전달함수 (relative transfer function), 채널 커플링 등으로 불린다.

상대전달함수의 추정을 위해서는 일반적인 최소평균자승오차 (MMSE) 추정법 대신 Shalvi 등이 제안한 비정적 특성을 이용한 방법 (nonstationarity-based method)을 이용하여 바이어스가 없는 추정 상대전달함수를 이용한다[10].

III. 제안된 주파수 영역 빔형성 기법

3.1. RTF-MVDR 알고리듬

TF-GSC에서 신호성분은 고정 빔형성기를 어떤 것을 사용하느냐에 따라 왜곡되는 정도가 달라지게 되는데, 주파수 영역 지향방향응답 (look-direction response)이 모든 주파수에 대해 1 (unity)인 경우 Gannot 등이 제안한 두 가지 고정 빔형성기에 의한 신호성분의 왜곡은 다음과 같다[3].

상대전달함수를 식 (7)로 정의하면 상대전달함수벡터는 식 (8)로 나타낼 수 있다.

$$\mathbf{H}(t,e^{j\omega}) = \left[H_1(t,e^{j\omega})\, H_2(t,e^{j\omega}) \cdots H_L(t,e^{j\omega})\right]^T \quad (8)$$

1) $\mathbf{W}_0(t,e^{j\omega}) = \dfrac{\mathbf{H}(t,e^{j\omega})}{\left\|\mathbf{H}(t,e^{j\omega})\right\|^2}$ 이면,

$$Y_{FBF,Signal}(t,e^{j\omega}) = \mathbf{A}_r(t,e^{j\omega})S(t,e^{j\omega}) \quad (9)$$

2) $\mathbf{W}_0(t,e^{j\omega}) = \mathbf{H}(t,e^{j\omega})$ 이면,

$$Y_{FBF,Signal}(t,e^{j\omega}) = \frac{\left\|\mathbf{A}(t,e^{j\omega})\right\|^2}{A_r^*(t,e^{j\omega})}S(t,e^{j\omega}) \quad (10)$$

이 된다[3]. 여기서, $\mathbf{W}_0(t,e^{j\omega})$는 고정 빔형성기의 계수벡터를 나타낸다.

식 (9)는 동등한 LCMV (Linearly Constrained Minimum Variance) 빔형성기의 제한조건(constraint)이

$$\mathbf{W}^H(t,e^{j\omega})\mathbf{A}(t,e^{j\omega}) = A_r(t,e^{j\omega}) \quad (11)$$

임을 의미한다. 즉, 식 (9)의 TF-GSC는 지향방향응답이 $A_r(t,e^{j\omega})$인 LCMV 빔형성기와 동일하다. 그러나, 전달함수벡터 $\mathbf{A}(t,e^{j\omega})$를 알 수 없다면 식 (11)의 제한조건을 갖는 LCMV 빔형성기는 구현이 불가능하지만, 아래와 같은 조작을 통해서 상대전달함수벡터를 알면 구현이 가능하도록 바꿀 수 있다.

식 (11)의 제한조건식은 결국 다음과 같다.

$$\mathbf{W}^H(t,e^{j\omega})\left(\frac{\mathbf{A}(t,e^{j\omega})}{A_r(t,e^{j\omega})}\Big|_j\right)=1$$

$$\therefore \mathbf{W}^H(t,e^{j\omega})\mathbf{H}(t,e^{j\omega})=1 \quad (12)$$

식 (12)의 제한조건을 만족하는 MVDR 빔형성기는, MVDR 빔형성기의 조향벡터로 $\mathbf{A}(t,e^{j\omega})$ 대신 $\mathbf{H}(t,e^{j\omega})$가 사용된 것과 같으므로, 식 (13)을 풀면 식 (14)의 계수벡터를 이 얻을 수 있다.

$$\min_{\mathbf{W}}\left\{\mathbf{W}^H(t,e^{j\omega})\mathbf{\Phi}_{\mathbf{XX}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right\}$$
$$subject\ to\ \mathbf{W}^H(t,e^{j\omega})\mathbf{H}(t,e^{j\omega})=1. \quad (13)$$

$$\mathbf{W}_{opt,RTF-MVDR}(t,e^{j\omega})=\frac{\mathbf{\Phi}_{\mathbf{XX}}^{-1}(t,e^{j\omega})\mathbf{H}(t,e^{j\omega})}{\mathbf{H}^H(t,e^{j\omega})\mathbf{\Phi}_{\mathbf{XX}}^{-1}(t,e^{j\omega})\mathbf{H}(t,e^{j\omega})} \quad (14)$$

식 (14)는 식 (6)과는 달리 신호원과 마이크간의 전달함수벡터 대신에, 상호전달함수벡터를 조향벡터로 사용한 것과 같다. 또한, TF-GSC와 마찬가지로 신호성분에 식 (11)에서처럼 소스와 기준 마이크 사이의 전달함수 $A_r(t,e^{j\omega})$만큼 왜곡이 생기므로 준최적 해로 볼 수 있다. 앞으로, 이 알고리듬을 RTF-MVDR 빔형성 기법으로 부르기로 한다.

3.2. RTF-DS (Delay-and-Sum) 알고리듬

신호와 마이크간의 전달함수를 알고 있다면, 식 (3)과 같은 주파수 영역 수신신호에 대한 최적의 Delay-and-Sum 빔형성기는 식 (15)과 같다.

$$\mathbf{W}_{opt,DS}(t,e^{j\omega})=\frac{\mathbf{A}(t,e^{j\omega})}{\mathbf{A}^H(t,e^{j\omega})\mathbf{A}(t,e^{j\omega})} \quad (15)$$

3.1절에서와 마찬가지로 전달함수 대신 상대전달함수를 조향벡터로 사용하면 식 (16)를 얻을 수 있고, 이 방법으로 필터링을 하면 신호성분이 RTF-MVDR 빔형성 기법과 마찬가지로 $A_r(t,e^{j\omega})$만큼 왜곡이 생긴다.

$$\mathbf{W}_{RTF-DS}(t,e^{j\omega})=\frac{\mathbf{H}(t,e^{j\omega})}{\mathbf{H}^H(t,e^{j\omega})\mathbf{H}(t,e^{j\omega})} \quad (16)$$

이 방법은 식 (9)에서 사용된 TF-GSC의 고정빔형성기를 빔형성기로 사용하는 것과 같으며, 앞으로 식 (16)과 같은 알고리듬을 RTF-DS 알고리듬으로 부르기로 한다.

IV. 실험 결과

4.1. 잡음 음성 데이터베이스

제안된 알고리듬의 성능을 평가하기 위하여, 자동차환경에서 잡음 음성 데이터베이스를 구축하였다.

그림 1과 같이, 2,000cc 중형승용차의 조수석 썬바이저에 5개의 무지향성 (omni-directional) 마이크를 6cm 간격으로 배치하고 프리앰프 및 디지털 멀티채널 레코더를 통하여 잡음과 깨끗한 음성을 따로 녹음하여 나중에 SNR별로 혼합된 잡음 음성을 만들 수 있도록 하였다 [11]. 가운데 위치한 3번 마이크가 조수석 헤드레스트 중심부와 일직선이 되도록 조정하고 조수석 화자가 정면을 바라보았을 때 음성이 거의 수직으로 입사하도록 어레이를 조정하였다.

잡음은 표 1과 같이 6가지 경우에 대해 녹음하였으며, 고속은 자동차 전용도로에서의 80 - 100 km/h의 속도를 말하고, 저속은 60km/h 이하의 신호대기가 포함된 시내주행을 말한다. 음악은 팝음악을 차량에 내장된 CD Player를 통해 재생했으며, 볼륨은 일반적인 상황과 최대한 비슷하도록 저속에서 창문을 닫았을 때는 작게, 고속에서 창문을 닫았을 때는 조금 더 크게, 저속에서 창문을 열었을 때에는 외부잡음을 감안하여 세 가지 경우 중에서 가장 크게 설정하였다. 또한, 저속에서 창문을 열었을 때에는 비정적인 외부 차량 잡음이 잘 유입되도록 되도록 주로 1차선에서 주행하였으며, 유턴차선에서 대기하면서 상대적으로 빠른속도로 진행하는 반대편 차량의 소음이 녹음되도록 하였다.

음성은 비교적 조용한 지하주차장에서 남자 17명, 여자 4명이 35단어를 단어 사이에 1초이상 쉬고 발음하도록 하였고, 그러한 녹음을 1인당 두 번씩 수행하였다.

그림 1. 실험 환경
Fig. 1. Experimental setup.

표 1. 주행 잡음 조건
Table 1. Condition of the car noise.

속도	창문	음악
High	Closed	Off
"	"	On
Low	"	Off
"	"	On
"	Open	Off
"	"	On

표 2. 인식실험에 사용된 파라메터
Table 2. Parameters used for recognition experiment.

window size	25ms
window shift rate	10ms
window type	Hamming
HMM type	Word model, DHMM
feature vector	M (12차 MFCC) + D(delta)M + DDM + E + DE + DDE
number of states	15
number of mixtures	8

녹음시 표본화율은 24kHz 였고, 녹음된 잡음과 음성으로부터 -5dB에서부터 2.5dB간격으로 10dB까지의 SNR을 갖는 잡음 음성 신호를 만들고 이를 8kHz의 표본화율로 다운샘플하였다. 이와 같은 처리는 화자별로 35단어씩 녹음된 단위별로 수행되었고, 나중에 HTK에서 인식 실험시 끝점 검출 오차의 영향을 없애기 위해 단어단위로 파일을 자를 수 있도록 수작업으로 단어의 시작점과 끝점의 시간을 기록하였다.

4.2. 인식 실험 및 결과

제안된 알고리듬의 성능 평가는 HTK를 이용한 인식률 비교를 통하여 이루어졌다. HTK의 특징추출 및 HMM 관련 파라메터는 표 2와 같다. 인식률 시험은 표본 데이터를 각각 남자 4명과 여자 1명으로 구성된 3개 세트와 남자 5명과 여자 1명으로 구성된 1개 세트, 총 4개의 배타적 세트로 나누어 round-robin 방식으로 시험하였다. 이 때, 훈련은 3번 마이크의 깨끗한 음성 신호로 하였고, 시험은 SNR별로 준비된 3번 마이크 잡음 음성 및 5가지 알고리듬으로 후처리된 신호로 하였다.

모든 알고리듬은 단구간 퓨리어 변환 영역에서 수행되었으며 이때 사용된 창함수는 256탭 hamming, FFT 크기는 512, 중첩은 75%로 하였다[12]. 성능비교를 위해 원격장 조향벡터를 이용한 far-DS (Delay-and-Sum), far-MVDR 및 상대전달함수를 조향벡터로 이용한 TF-GSC, RTF-MVDR, RTF-DS 알고리듬을 적용하였다.

상대전달함수의 추정은 잡음레벨이 커질수록 부정확해지므로, 본 논문에서는 상대전달함수의 오차로 인한 영향을 최소화할 수 있도록 화자별로 35단어씩 녹음된 모든 단위 음성에 대해 각각 깨끗한 음성으로부터 비정적 방법을 이용하여 구한 값을 모든 잡음 음성에 사용하였다[10,13]. 상대전달함수의 추정에 사용된 깨끗한 음성 신호의 녹음은 조수석에만 사람이 있는 상태에서 녹음되었고, 잡음신호는 운전석에만 사람이 있는 상태에서 녹음되었기 때문에, 이렇게 추정된 상대 전달 함수를 인위적으로 합산하여 만든 잡음 음성 신호에 적용할 경우, 추가적인 추정 오차가 있을 수는 있지만 상대전달함수는 비교적 가까운 거리에 있는 마이크 간의 전달함수임을 감안할 때 직접 경로를 통한 영향이 훨씬 클 것이므로 큰 문제는 없을 것으로 판단된다.

자동차 환경에서 간섭 신호의 시간적 특성은 시간에 따라 변화할 수 있지만 공간적 특성은 외부적인 요인이 없는 한 크게 변화하지 않는다고 가정하고, far-MVDR 및 RTF-MVDR 알고리듬의 계수 갱신은 35단어씩 녹음된 단위 음성에 대해 잡음만 있는 앞부분의 약 600ms 구간에서 추정한 공간상관행렬로부터 한 번만 구해서 전체 35단어 단위 음성에 적용하였다. TF-GSC를 제외한 나머지 알고리듬들은 계수를 한번만 계산해서 35문장으로 이루어진 단위 음성에 대해 같은 계수로 필터링을 하므로, 계수 계산에 필요한 계산량을 무시하고 필터링에 드는 계산량만 따진다면 L탭의 복소 필터링만 필요하지만, TF-GSC는 시변하는 필터계수에 의한 시간영역 에일리어징 (time-domain aliasing)에 의한 잡음을 막기 위한 IFFT와 계수절삭 및 FFT 과정을 빼고도 약 5배의 계산이 필요하므로 제안된 알고리듬이 약 5배 이상의 계산상의 이득이 있다[3,14].

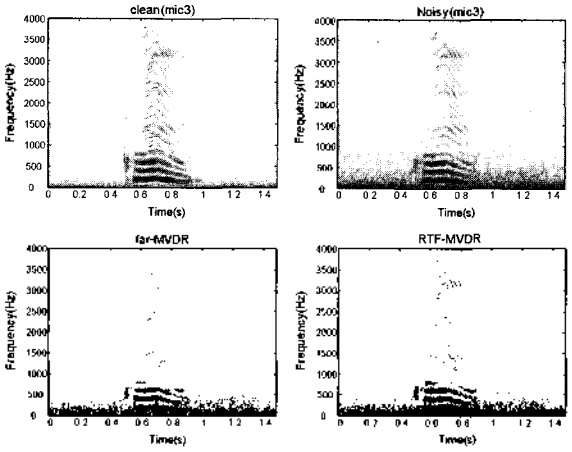

그림 2. far MVDR과 RTF MVDR의 비교 예 "고속, 창문닫고, 음악끔", SNR=0dB, '공'
Fig. 2. Example: Comparison of RTF MVDR with far MVDR "high speed, window closed, music off", SNR=0dB, '공'.

그림 2는 far-MVDR과 RTF-MVDR 빔형성기 출력신호의 스펙트로그램을 나타낸 것이다. far-MVDR이 RTF-MVDR에 비해 잡음은 더 많이 제거하지만, 근접장 효과 때문에 신호까지 감쇄되며 특히 1kHz 이하의 저주파 부분이 많이 감쇄되는 것을 볼 수 있다.

HTK를 이용한 인식률을 비교한 그래프를 그림 3에서 그림 8에 보였다. 기준이 될 수 있는 잡음이 없는 깨끗한 신호로 테스트한 결과는 94.63%였다.

먼저 far-MVDR과 RTF-MVDR의 성능을 비교해보면, 전반적으로 RTF-MVDR이 높은 인식률을 보였으며, SNR이 낮을수록 차이가 심했고 최대 15%정도 차이가 났다. far-MVDR의 인식률이 낮은 이유는 근접장 효과에 의한 신호 감쇄 및 왜곡 때문이라고 판단되며, "저속, 창문닫고" 일 때에는 음악 켬/끔에 상관없이 잡음신호보다도 낮은 인식률을 나타냈는데, 그 이유는 "저속, 창문닫고" 일 때에는 주로 엔진룸쪽에서 유입되는 극저주파 대역에 에너지가 집중된 엔진 노이즈가 강하기 때문에 앞과 뒤를 구분 못하는 원격장 선형 어레이의 특성상 목적 신호 주변에서 간섭신호가 들어오는 경우와 같으므로 저주파 대역에서 목적 신호의 감쇄 및 왜곡이 심하기 때문이라고 판단된다.

far-DS와 RTF-DS의 경우에는 "저속, 창문열고" 일 때는 비슷한 인식률을 보였고, 그 이외의 상황에서는 RTF-DS가 최대 3% 정도 높은 인식률을 보였다. Far-MVDR과 RTF-MVDR의 인식률 비교에 비해 차이가 적은 이유는 Delay-and-Sum 빔형성기는 간섭신호 방향으로 널링을 하지 않으므로, 단순히 RTF-DS가 시간지연보상을 더 충실히 한 효과밖에 없기 때문이라고 판단된다.

위의 결과로 미루어보아, 상대전달함수벡터를 조향벡터로 이용한 제안한 방법이 원격장 조향벡터를 사용한 방법보다 인식성능을 향상시킴을 알 수 있다.

마지막으로, TF-GSC와 고정된 계수로 필터링 하는 부분만 생각했을 때 TF-GSC에 비해 상대적으로 계산량이 적은 RTF-MVDR의 인식률을 비교해보면, "저속, 창문열고, 음악켬" 일 때에는 최대 7% 차이로 RTF-MVDR이 저조한 인식률을 보였다. 그러나, 그 이외의 경우에서는 최대 2.5% 내에서 RTF-MVDR과 TF-GSC의 성능

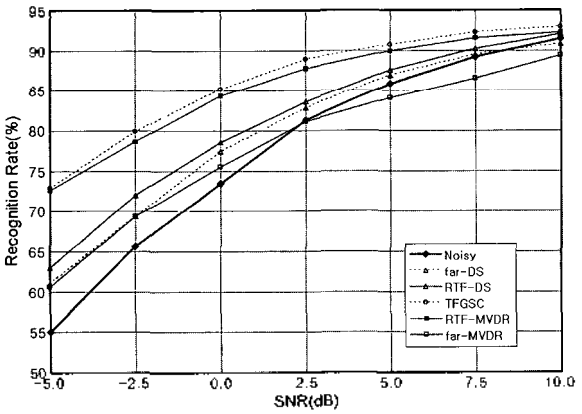

그림 3. 인식률 (고속, 창문닫고, 음악끔)
Fig. 3. Recognition rate (high speed, window closed, music off).

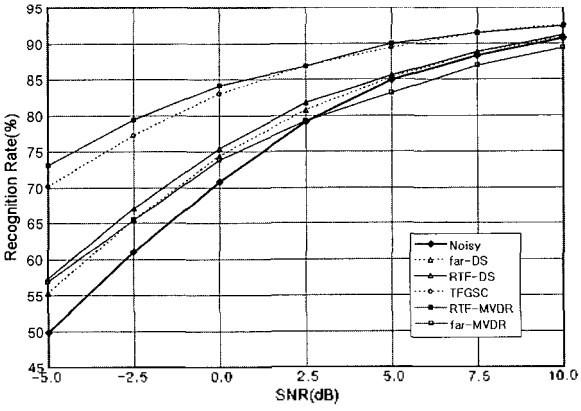

그림 4. 인식률 (고속, 창문닫고, 음악켬)
Fig. 4. Recognition rate (high speed, window closed, music on).

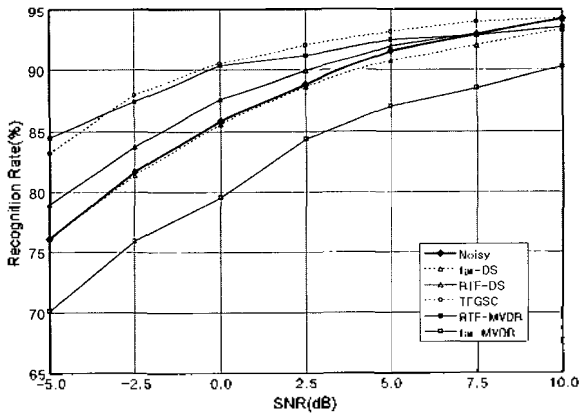

그림 5. 인식률 (저속, 창문닫고, 음악끔)
Fig. 5. Recognition rate (low speed, window closed, music off).

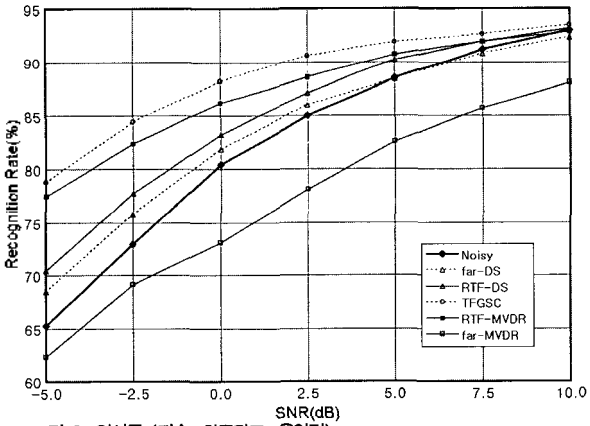

그림 6. 인식률 (저속, 창문닫고, 음악켬)
Fig. 6. Recognition rate (low speed, window closed, music on).

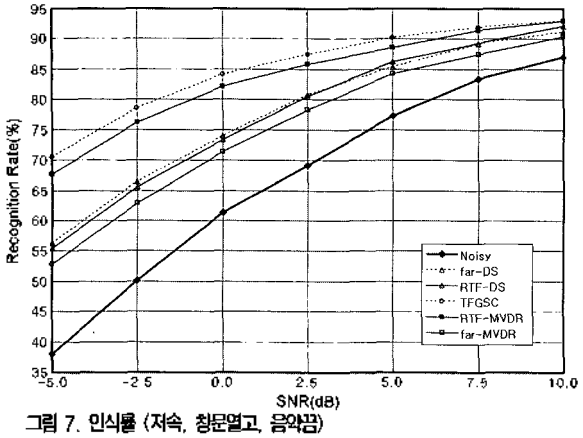

그림 7. 인식률 (저속, 창문열고, 음악끔)
Fig. 7. Recognition rate (low speed, window opened, music off).

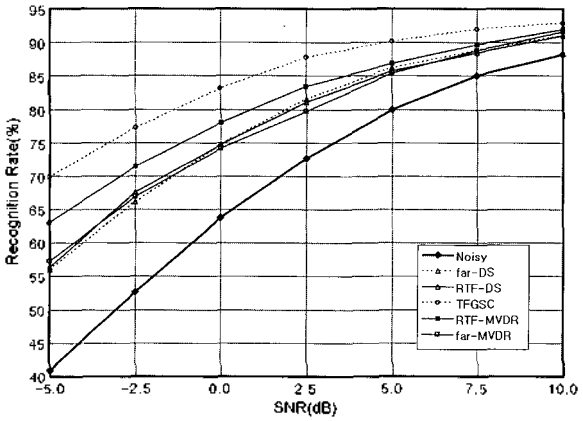

그림 8. 인식률 (저속, 창문열고, 음악켬)
Fig. 8. Recognition rate (low speed, window opened, music on).

은 대등하였다.

"저속, 창문열고, 음악켬" 일 때에만 유독 RTF-MVDR의 성능이 저조한 이유는 "저속, 창문열고, 음악켬" 일 때가 유턴 구간에서 반대편에서 마주오는 방향으로 빠르게 진행하는 차량에 의한 비정적인 외부 차량 잡음이 가장 빈번했고, 오디오 볼륨도 가장 높았기 때문에 필터 계수를 초기에 한 번만 갱신하는 방법의 한계로 판단된다.

V. 결 론

본 논문에서는 상대전달함수를 조향 벡터로 사용하는 RTF-MVDR 알고리듬과 RTF-DS 알고리듬을 제안하였고, 자동차 환경에서 기존 알고리듬과의 인식률 시험을 통해 이들의 성능을 평가하였다.

제안된 알고리듬은 주변 환경의 변화가 적은 충분한 시간동안에 초기에 한 번 계산된 계수를 그대로 사용하도록 하였다. 이에, RTF-DS 알고리듬은 기존의 원격장 조향벡터를 사용하는 Delay-and-Sum 알고리듬과 같은 계산량으로 최대 3%정도 인식률 향상이 있었고, RTF-MVDR 알고리듬도 기존의 원격장 조향벡터를 사용하는 MVDR 알고리듬과 같은 계산량으로 최대 15%의 인식률 향상이 있었다.

제안된 알고리듬은 자동차 환경 뿐만 아니라, 임의의 지오메트리를 갖는 센서 어레이에 적용가능하며, 추적능력이 있고 잡음환경하에서 적은 추정오차를 갖는 상대전달 함수 추정 알고리듬과 결합한다면 전달함수의 시간에 대한 변화가 더 큰 환경에도 적용이 가능하다.

참 고 문 헌

1. M. Bransdstein and D. Ward, *Microphone Arrays*, (Springer, 2001).
2. H. L. Van Trees, *Optimum Array Processing*, (Wiley-Interscience, 2002).
3. S. Gannot, D Burnshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech", IEEE Trans. Signal Processing, **49**, 1614-1626, Aug. 2001.
4. F. Asano, H. Aso, and T. Matsui, "Sound Source Localization and Separation in Near Field", IEICE Trans. Fundamentals, E38-A, 2286-2294, Nov. 2000.
5. F. Asano et al., "Real-time sound source localization and separation system and its application to automatic speech recognition", in Proc. EUROSPEECH2001, Aalborg, Denmark, Sept. 2001, 1013-1016.
6. S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech", IEEE Trans. Speech and Audio Processing, **5**, 425-437, Sept. 1997.
7. O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing", Proc. IEEE, **60**, 926-935, Jan. 1972.
8. L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming", IEEE Trans. Antennas and Propagations, **AP-30**, 27-34, Jan. 1982.
9. http://htk.eng.cam.ac.uk/
10. O. Shalvi and E. Weinstein, "System identification using nonstationary signals", IEEE Trans. Signal Processing, **44**, 2055-2063, Aug. 1996.
11. I. Cohen, "Multichannel Post-Filtering in Nonstationary Noise Environments", IEEE Trans. Signal Processing, **52**, 1249-1150, May 2004.
12. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, (Prentice-Hall, 1978).
13. I. Cohen, "Relative transfer function identification using speech signals", IEEE Trans. Speech and Audio Processing, **12**, 451-459, Sept. 2004.
14. R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, (Prentice-Hall, 1983).

저자 약력

• 한 철 희 (Chul-Hee Han)

1997년 2월: 중앙대학교 공과대학 전자공학과 졸업
1999년 2월: 연세대학교 공과대학 전자공학과 석사
2000년 3월-현재: 연세대학교 공과대학 전기전자공학과 박사과정
※주관심분야: 디지털 신호처리, 적응 신호처리, 음성 신호처리, 마이크로폰 어레이

• 강 홍 구 (Hong-Goo Kang)

1989년 2월: 연세대학교 전자공학과 졸업
1991년 2월: 연세대학교 전자공학과 석사
1995년 2월: 연세대학교 전자공학과 박사
1996년~2002년: Senior Technical Staff Member of AT&T Labs-Research
2002년 - 현재: 연세대학교 전기전자공학과 교수

• 황 영 수 (Youngsoo Hwang)

1982년 2월: 연세대학교 전자공학과 졸업
1984년 2월: 연세대학교 전자공학과 석사
1990년 2월: 연세대학교 전자공학과 박사
1989년~현재: 관동대학교 전자정보통신기술공학부 교수
1991년 1월~2000년 2월: OGI CSLU in USA Visiting Scholar
※주관심분야: 음성 신호처리, 음향학, 멀티미디어

• 윤 대 희 (Dae-Hee Youn)

한국음향학회지, 제24권 제3호 참조
현재: 연세대학교 전기전자공학과 교수