한국음향학회지 제33권 제1호 pp. 54~59 (2014)
The Journal of the Acoustical Society of Korea Vol.33, No.1 (2014)
http://dx.doi.org/10.7776/ASK.2014.33.1.054

pISSN : 1225-4428
eISSN : 2287-3775

CASA 기반의 마이크간 전달함수 비 추정 알고리즘

CASA Based Approach to Estimate Acoustic Transfer Function Ratios

신민규, 고한석†

(Minkyu Shin and Hanseok Ko†)

고려대학교 전자전기전파공학부

(접수일자: 2013년 9월 24일; 수정일자: 2013년 10월 25일; 채택일자: 2013년 10월 25일)

초 록: 본 논문은 비정상(nonstationary)특성을 가지는 잡음환경에서 마이크간 전달함수 비 (RTF, Relative Transfer Function) 추정 알고리즘을 제안한다. 음성을 이용한 다양한 기기에 다중 마이크를 이용한 잡음제거 기술은 널리 사용되며, 이때 각 마이크간의 입력 신호 사이의 관계는 필수적으로 추정되어야 한다. 본 논문에서는 기존의 OM-LSA (Optimally-Modified Log-Spectral Amplitude)기반의 추정 방식에 CASA (Computational Auditory Scene Analysis)를 접목시킨 방식을 제안한다. 제안한 방법의 성능 검증을 위하여 비정상 백색 잡음 (nonstationary white Gaussian noise) 환경에서 10명 화자 발음을 이용한 마이크간 전달함수 비 추정 성능 평가 실험을 수행하였다. 잡음 신호가 초당 8dB 증감하는 환경에서 SBF (Signal Blocking Factor)가 평균 2.65dB 개선됨을 확인하였다.
핵심용어: 전달함수 차이, 마이크 어레이, 잡음제거

ABSTRACT: Identification of RTF (Relative Transfer Function) between sensors is essential to multichannel speech enhancement system. In this paper, we present an approach for estimating the relative transfer function of speech signal. This method adapts a CASA (Computational Auditory Scene Analysis) technique to the conventional OM-LSA (Optimally-Modified Log-Spectral Amplitude) based approach. Evaluation of the proposed approach is performed under simulated stationary and nonstationary WGN (White Gaussian Noise). Experimental results confirm advantages of the proposed approach.
Keywords: System identification, Relative transfer function, Multi-microphone speech enhancement
PACS numbers: 43.60. Fg

I. 서 론

멀티 마이크를 활용함에 있어서, 동일한 목표 신호를 관측하더라도, 각 마이크에 입력된 신호들 사이에 차이가 존재한다. 이 차이를 단순히 시간지연으로 모델링 하는 경우도 있지만,[1] 실제 환경에서 이 같은 모델링은 상당한 성능 하락을 유발 한다.[2] 목표신호음원의 위치로부터 각 마이크까지의 전달함수에는 시간 지연 외에도 반향, 마이크 특성 등의 다양한 차이가 존재하기 때문이다. 이 전달함수들의 비를 마이크간 전달함수 비라고 한다. 음성을 이용한 다양한 기기에서, 멀티 마이크를 이용한 잡음제거 기술은 널리 사용되며, 이때 마이크간 전달함수 비는 필수적으로 추정되어야 한다. 예를 들어 LCMV(Linearly Constrained Minimum Variance) 방식의 잡음제거 중 가장 널리 사용되는 GSC(Generalized Sidelobe Canceller)[1]는 blocking matrix 를 이용하여 noise reference signal을 얻는다. 만약 마이크간 전달함수 비 추정이 제대로 되지 않는다면, blocking matrix 의 출력에 목표 신호가 남아 있게 된다. 이로 인해, blocking matrix 뒤에 오는 noise canceller 가 잡음뿐만 아니라 목표 신호도 제거하게 되어 성능이 하락한다. 처음 GSC가 제안되었을 때,[1] 각 마이크에

†**Corresponding author:** Hanseok Ko (hsko@korea.ac.kr)
Engineering building room 419 or room 439 Department of Electronics and Computer Engineering, Korea University Anam Campus, Anam-Dong 5-Ga, Seongbuk-Gu, Seoul 136-713, Republic of Korea
(Tel: 82-2-3290-3239, Fax: 82-)

입력된 목표 신호들의 차이로써 오로지 시간지연만 존재한다고 가정 하였다. Gannot[3]은 각 마이크에 입력되는 목표 신호들 사이에 임의 전달 함수(arbitrary transfer function) 의 존재를 모델링 하였으며, 이 임의 전달 함수를 추정하기 위해 추정 단계(system identification step) 를 기존 GSC에 추가한 TF-GSC(general Transfer Function GSC)를 제안하였다. 그가 제안한 TF-GSC에서 추정 단계 는 목표신호의 비정상성을 활용하여 수행 되었다.[4] 이후, Cohen[5-6]은 이 비정상성 기반 방식(nonstationarity-based method) 의 단점을 보완하기 위해 OM-LSA[7] 기반의 마이크간 전달함수 비 추정 방식을 제안하였다. OM-LSA기반의 마이크간 전달함수 비 추정 방식에서는 목표 신호의 존재여부에 대한 지시 함수를 OM-LSA 과정중 계산되는 음성 존재 확률(speech presence probability)로부터 추정하였으며, 추정된 지시 함수에 따라 입력 신호의 시간-주파수 구역(time-frequency bin) 을 선택적으로 활용하였다.

기존 OM-LSA기반의 마이크간 전달함수 비 추정 방식은 잡음환경이 비정상 특성을 갖는 경우 마이크간 전달함수 비 추정에 사용 할 시간-주파수 구역 선택에 오차가 생기는 단점이 있다. 이는 주파수 각 대역의 기존 파워값을 바탕으로 잡음을 추정하기 때문에 기존 파워값과 상이한 입력신호가 입력될 경우 해당 입력 성분을 목표 신호로 간주하기 때문이다. 본 논문에서는 이러한 단점을 보완하기 위한 CASA 기반의 마이크간 전달함수 비 추정 알고리즘을 제안한다. 제안하는 방법은 목표 신호의 존재여부에 대한 지시 함수 를 구하는데 있어서 기존 방식과 더불어, 음조(pitch), 시간-주파수 영역 상의 인접한 정도, 진폭 변조 등의 정보를 활용하기 위해 CASA이론에 기반한 IBM(Ideal Binary Mask)을 활용한다.

이어지는 섹션2 에서는 모델링 및 기존의 마이크간 전달함수 비 추정 방식에 대한 설명한다. 섹션 3에서는 제안되는 알고리즘을 설명하고, 섹션 4에서는 실험을 진행한다. 이후 섹션 5에서 결론을 맺는다.

II. Conventional Algorithms

먼저, 다음과 같은 모델링을 수행한다. $x(t)$와 $y(t)$를 각각 1번 마이크와 2번 마이크의 입력 신호라고 한다.

$$x(t) = s(t) + u(t), \tag{1}$$

$$y(t) = a(t) \circ s(t) + w(t). \tag{2}$$

식(1)의 $s(t)$ 는 1번 마이크에 입력된 목표신호이고, 식(2)의 $a(t)$ 는 1번 마이크와 2번 마이크 에 입력된 목표신호사이의 마이크간 전달함수 비이다. $u(t)$와 $w(t)$는 각각의 마이크에 입력된 잡음신호이다. 여기서 $\circ$ 는 컨벌루션을 나타낸다. 식(1), (2)를 선형 시간 시스템으로 표현하면 식(3), (4)와 같다

$$y(t) = a \circ x(t) + v(t), \tag{3}$$

$$v(t) = w(t) - a \circ u(t). \tag{4}$$

y 와 x 사이의 cross-PSD(Power Spectral Density) 는 아래와 같이 표현된다.

$$S_{yx}(k,l) = A(k,l)S_{xx}(k,l) + S_{vx}(k,l). \tag{5}$$

여기서 k 는 주파수 인덱스 이며, l 은 프레임 인덱스 이다. 우리는 $s(t)$와 $u(t)$, $w(t)$가 상관 관계가 없다고 가정 하며, 따라서 cross-PSD 는 아래와 같이 표현된다.

$$S_{yx}(k,l) = A(k,l)S_{ss}(k,l) + S_{wu}(k,l). \tag{6}$$

각각의 항목들을 추정 하면, 식은 아래와 같이 표현된다.

$$\hat{S}_{yx}(k,l) - \hat{S}_{wu}(k,l) = A(k)\hat{S}_{ss}(k,l) + \epsilon(k,l). \tag{7}$$

여기서 $\epsilon(k,l)$는 추정 오차 이다.
식(7) 을 행렬로 쓰면 식(8) 과 같다

$$\begin{aligned}\hat{\psi}(k) &\equiv \begin{bmatrix} \hat{S}_{yx}(k,1) - \hat{S}_{wu}(k,1) \\ \hat{S}_{yx}(k,2) - \hat{S}_{wu}(k,2) \\ \vdots \\ \hat{S}_{yx}(k,L) - \hat{S}_{wu}(k,L) \end{bmatrix} \\ &= \begin{bmatrix} \hat{S}_{ss}(k,1) \\ \hat{S}_{ss}(k,2) \\ \vdots \\ \hat{S}_{ss}(k,L) \end{bmatrix} A(k) + \begin{bmatrix} \epsilon(k,1) \\ \epsilon(k,2) \\ \vdots \\ \epsilon(k,L) \end{bmatrix} \\ &\equiv \hat{S}_{ss}(k)A(k) + \epsilon(k). \end{aligned} \tag{8}$$

각 마이크에 입력된 목표신호에 대한 전달함수의 차이인 $A(k)$ 를 추정하기 위해서는, 추정에 사용되는 시간 프레임들이 음성 신호를 포함하고 있어야 한다. 따라서 음성신호의 존재 여부를 표현하는 지시함수(indicator function) $I(k,l)$ 를 정의하고, $I(k,l)$ 값이 1일 경우에만 해당 k,l 값에 해당하는 시간-주파수 구역을 $A(k)$추정에 사용한다. $I(k,l)$는 $S_{ss}(k,l) \neq 0$ 일 때 1의 값을 가지며, 그 외의 경우에는 0의 값을 갖는다. 이러한 $I(k,l)$를 대각행렬 형태로 표현한 것이 $I(k)$이며, $I(k)$는 그 대각 성분으로 $I(k,1), I(k,2), ..., I(k,L)$을 갖는다. 가중치 최소 자승법 방식으로 $A(k)$ 를 추정하면 식(9)와 같으며

$$\begin{aligned}&\hat{A}(k)\\&= \arg min_{A(k)}[I(k)\epsilon(k)]^H W[I(k)\epsilon(k)]\\&= \arg min_{A(k)}\\&\quad [\hat{\psi}(k)-\widehat{S_{ss}}(k)A(k)]^H IWI[\hat{\psi}(k)-\widehat{S_{ss}}(k)A(k)]\\&= [\widehat{S_{ss}^T}(k)I(k)\,WI(k)\widehat{S_{ss}}(k)]^{-1}\widehat{S_{ss}^T}(k)I(k)\,WI(k)\hat{\psi}(k)\end{aligned} \tag{9}$$

식(9)에서 추정된 $A(k)$의 분산을 최소화 하는 $I(k)\,WI(k)$ 는 식(10)과 같다.

$$I(k)\,WI(k) = I(k)[cov(\epsilon(k))]^{-1}I(k). \tag{10}$$

식(10)에서 $cov(\epsilon)$는 아래 식(11)과 같이 계산할 수 있다.

$$cov(\epsilon(k,l),\epsilon(k,l)') = \begin{cases} S_{ss}(k,l)S_{vv}(k,l), & \text{if } l=l' \\ 0 & else \end{cases}. \tag{11}$$

식(10)을 식(9)에 입력하면 아래 식(12)와 같다.

$$\hat{A}= [\hat{S}_{ss}^T I[cov(\epsilon)]^{-1}I\hat{S}_{ss}]^{-1}\hat{S}_{ss}^T I[cov(\epsilon)]^{-1}I\hat{S}_{ss}\hat{\psi}. \tag{12}$$

이때, 추정된 A 의 분산은 식(13)과 같다.

$$var\hat{A}= (\hat{S}_{ss}^T I[cov(\epsilon)]^{-1}I\hat{S}_{ss})^{-1}. \tag{13}$$

OM-LSA기반의 마이크간 전달함수 비 추정 방식에서는 잡음 신호를 정적잡음으로 가정 한다. 이 가정은 목표신호가 음성일 경우, 음성에 비해 변화속도가 느린 잡음환경을 반영한다. 평균연산 을 식(14)와 같이 정의한다.

$$< \phi(k,l) >_l \equiv \frac{1}{L}\sum_{l=1}^{L}\phi(k,l). \tag{14}$$

식(11)를 식(12)에 대입하고, 식(14)를 이용해 표현하면 식(15)와 같이 $\hat{A}$ 를 표현할 수 있다.

$$\hat{A}= \frac{< I(k,l)[\hat{S}_{yx}(k,l)-\hat{S}_{wu}(k,l)] >_l}{< I(k,l)\hat{S}_{ss}(k,l) >_l}. \tag{15}$$

III. 제안하는 알고리즘

OM-LSA기반의 마이크간 전달함수 비 추정 방식에서는, $I(k,l)$를 구하기 위해 OM-LSA 과정에서 계산된 음성 존재 확률을 사용하며, 그 과정은 식(16)과 같다.

$$\hat{I}(k,l) = \begin{cases} 1 & \text{if } p(k,l) \geq p_0 \\ 0 & otherwise \end{cases}. \tag{16}$$

여기서 $p(k,l)$은 OM-LSA 방식을 통해 얻을 수 있는 음성 존재 확률이고, p_0는 미리 정해진 기준 값이다. 만약 $I(k,l)$의 추정에 오차가 발생하여 음성신호 가 존재하지 않는 시간-주파수 구역을 음성신호가 존재한다고 판단하였다면(false alarm), 이로 인해 $\hat{A}$는 잘못된 값으로 추정될 것이다. 또한, 음성신호가 존재하는 시간-주파수 구역을 존재하지 않는다고 판단한다면(miss), 추정에 사용할 수 있는 시간-주파수 구역이 줄어들 것이고 이로 인해 더 많은 입력신호를 사용해야 한다. p_0는 위 두 가지 오차(false alarm 과 miss)의 교환 관계를 조절한다.

잡음신호에 harmonicity가 없을 때, 음성신호의 harmonicity 는 음성의 존재 여부에 대한 추가적인 정보로 활용이 가능하며, 본 논문에서는 이러한 추가적인 정보를 활용하기 위한 방법으로 CASA 알고리즘 중 Wang[8-9]이 제안한 방식을 적용한다. ASA (Auditory Scene Analysis) 이론에 근거하여, Wang은 음성중 유성음에 대한 잡음제거 알고리즘을 제안하였으며, 그 원리는 다음과 같다. 우선 시간에서 연속적이면서 cross-channel correlation 또는 진폭 변조의

상관(common amplitude modulation) 값이 일정 수준 이상인 시간-주파수 구역들을 모아서 구획(segment)들을 생성한다. 이렇게 구성된 구획들 중 음조 가 비슷한 구획 들만 선택하여 그룹을 생성하고 한 그룹 내에 있는 구획 들은 하나의 음원으로부터 발생했다고 가정한다. 이후, 그룹들 중, 음성의 음조(예: dominant pitch) 에 대응하는 그룹만을 복원하는 방법으로 잡음이 제거되는데, 이때 복원에 사용될 시간-주파수 구역들을 나타내는 마스크가 IBM 이다.

IBM 은 OM-LSA기반의 마이크간 전달함수 비 추정 방식에서 사용되었던 지시 함수와 동일한 역할을 수행 하지만 음성의 음조 , 시간-주파수 영역상의 인접한 정도, 진폭 변조 등의 정보를 활용하므로, 기존 방식의 지시 함수가 오류를 범하기 쉬운 비정상 특성의 잡음환경에서 강인한 성능을 보인다.[8] 따라서, OM-LSA기반의 마이크간 전달함수 비 추정 방식에서 사용되었던 지시 함수와 CASA 기반의 IBM을 동시에 활용한 새로운 지시 함수로 기존의지시 함수 $I(k,l)$를 대체할 경우, 비정상 특성을 갖는 잡음환경에서 성능을 개선을 기대할 수 있다. 제안하는 방식에서 사용할 지시 함수는 식(17)과 같이 정의 된다.

$$\hat{I}_{proposed}(k,l)=\begin{cases}1 & \text{if } \hat{I}_B(k,l)\times\hat{I}_C(k,l)\neq 0\\ 0 & otherwise\end{cases}. \quad (17)$$

여기서 $I_B(k,l)$는 OM-LSA기반의 마이크간 전달함수 비 추정 방식에서 사용하였던 지시 함수이며, $I_C(k,l)$는 IBM을 지시 함수로 표현한 것이다. 따라서 두가지 지시 함수가 모두 1 일 때에만 해당 시간-주파수 구역이 마이크간 전달함수 비의 추정에 사용된다.

IV. 실험 결과

이번 섹션에서 OM-LSA기반의 마이크간 전달함수 비 추정 방식과 제안한 방식의 성능을 비교한다. 잡음 신호는 정상 백색 잡음과 비정상 백색 잡음이 사용된다. 비정상 백색 잡음은 초당 6-8 dB 만큼 2초 동안 증가하고, 같은 비율로 2초 동안 감소하는 잡음을 생성하여 사용한다. 음성신호는 DARPA RM data base[10] 의 훈련용 데이터베이스 중 여성 5명, 남성 5명의 음성을 사용하였다. 음성의 길이는 각 화자당 20 s가 되도록 정렬하였으며, 8 kHz 로 셈플링 되었다. 성능평가기준으로는[5]에 사용된 것과 동일한 방법인 SBF를 이용하였으며, 그 정의는 식(18)과 같다.

$$SBF=10\log_{10}\frac{E\{s^2(t)\}}{E\{r^2(t)\}}\,[dB]. \quad (18)$$

$E\{s^2(t)\}$는 음성 신호(clean)의 에너지 이며, $E\{r^2(t)\}$은 누출 신호의 에너지 이다. 누출 신호는 식(19)와 같이 정의 되며, 마이크 2 번에 입력된 음성신호와, 마이크 2번에 입력된 신호를 1번에 입력된 음성신호를 이용하여 추정한 신호 사이의 차를 의미한다.

$$r(t)=a(t)\circ s(t)-\hat{a}(t)\circ s(t). \quad (19)$$

2번 마이크에 입력된 음성신호에 적용된 전달함수는 식(20)과 같다.

$$a(t)=\delta(t-6T)-0.5\delta(t-7T)+0.25\delta(t-8T). \quad (20)$$

여기서 T는 셈플링 주기($125\mu s$) 이다. 2번 마이크에 입력된 노이즈는 1번 마이크에 입력된 잡음 $u(t)$ 를 이용하여 다음과 같이 생성되었다.

$$w(t)=g(t)\circ u(t), \quad (21)$$

$$g(t)=-\delta(t)-0.5\delta(t-T)+0.1\delta(t-2T). \quad (22)$$

Table 1은 0 ~10 dB SNR 의 정상 백색 잡음 환경에서의 기존 방식과 제안한 방식의 성능(SBF)을 나타낸다. 성능 비교를 위해 10명의 화자에 대한 결과의 평균을 사용하였다. 제안하는 알고리즘의 경우, OM-LSA기반의 마이크간 전달함수 비 추정 방식의 지시 함수 와 CASA알고리즘을 통해 계산된 IBM 모두를 만족하는 시간-주파수 구역만을 사용하므로, 사용되는 시간-주파수 구역의 개수를 충분히 확보하기 위해 지시 함수의 기준값인 p_0 값이, 비교 대상이 되는 기존 알고리즘의 p_0 비해 낮게 설정 되었다.

정상잡음의 경우 제안한 방식과 기존 방식의 성능에 큰 차이가 없음을 볼 수 있다. 입력 SNR 이 0 dB 인

Table 1. Experiment result in stationary noise (SBF).

Input Noise (dBSNR)	Conventional $p_0 = 0.8$ (dB)	Proposed $p_0 = 0.7$ (dB)	Improvement (dB)
0	-0.61	0.78	1.38
5	1.19	1.92	0.72
10	4.77	4.72	-0.05

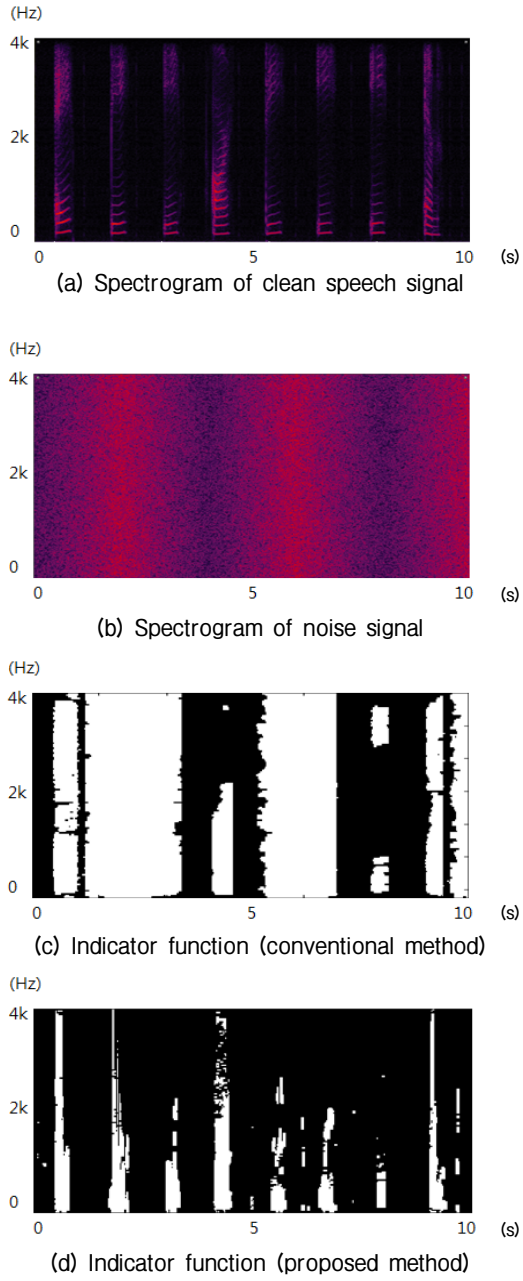
Table 2. Experiment result in nonstationary noise (SBF).

Increase (dB/s)	Input Noise (dBSNR)	Conventional $p_0 = 0.8$ (dB)	Proposed $p_0 = 0.7$ (dB)	Improvement (dB)
6	0	1.85	3.75	1.90
	5	5.29	6.56	1.26
	10	8.53	9.71	1.18
7	0	2.31	4.58	2.28
	5	5.54	6.99	1.45
	10	8.62	11.23	2.61
8	0	3.29	5.83	2.54
	5	6.22	8.45	2.23
	10	9.52	12.69	3.17

경우에 제안한 방식을 사용한 결과가 1.38 dB SNR 개선 되었으나, 입력 SNR 이 5, 10 dB 인 경우, 성능의 차이가 1 dB 이하의 수준이다. 잡음이 정상 잡음 이고, 입력 SNR 이 일정 수준 이상으로 보장되면, 기존의 방식에서의 지시 함수가 만족할 만한 성능을 보이기 때문으로 분석된다. 비정상 특성을 갖는 잡음을 이용한 실험의 결과는 Table 2와 같다.

잡음이 비정상 특성을 갖는 환경에서, 제안된 방식이 개선된 성능을 보인다. 특히, 잡음 환경의 비정상 특성정도 가 증가할수록 개선정도 역시 증가한다. 제안한 알고리즘의 성능 개선 원리를 분석하기 위해 시간-주파수 영역에서 기존 방식과 제안하는 방식으로 추정한 지시 함수 $I(k,l)$를 비교해 보면 Fig. 1과 같다.

Fig. 1 중 (a), (b) 는 각각 음성신호(clean speech, 8글자의 알파벳을 발화)와 잡음신호(8dB/s 로 증감하는 백색 잡음)의 스펙트로그램이다. 입력된 신호(noisy speech, 0dB SNR)로부터, 기존 방식을 이용하여 지시 함수를 구한 결과가 (c)이며 제안하는 알고리즘으로 지시 함수를 구한 결과가 (d) 다. 흰색은 1의 값을, 검은색은 0의 값을 표현한다. 기존 방식은 잡음의 크기



(a) Spectrogram of clean speech signal

(b) Spectrogram of noise signal

(c) Indicator function (conventional method)

(d) Indicator function (proposed method)

Fig. 1. Spectrogram of input signal and corresponding indicator function.

가 빠르게 변화하는 시간-주파수 구역들에 대해 음성 신호로 인해 입력신호가 크게 변화하는 것으로 오인하여 음성신호가 존재한다고 판단하는 경우가 빈번하게 발생한다. 반면, 제안하는 방식은 음성 고유의 특성인 harmonicity 의 존재여부를 이용하므로, 이런 오류를 발생시키지 않는 것을 볼 수 있다.

V. 결 론

기존 OM-LSA기반의 마이크간 전달함수 비 추정 방식에서 제안된 알고리즘은 비정상성 기반 방식 보다 성능이 뛰어났으며, 특히 비정상 특성을 갖는 잡음에서 개선된 성능을 보였다. 하지만 OM-LSA기반의 마이크간 전달함수 비 추정 방식 역시 잡음의 비정상 특성 정도가 심해지면 성능은 하락한다. 이 부분을 보완하기 위해, 본 논문에서는 음성 고유의 특징을 이용하여 주변 소리와 음성을 구분하는 CASA 알고리즘을 OM-LSA기반의 마이크간 전달함수 비 추정 방식에 적용하는 방법을 제안하였다. 실험을 통해 잡음 신호가 초당 8 dB 만큼 증감 하는 환경에서 마이크간 전달함수 비 추정 성능이 SBF 기준 평균 2.65 dB 개선되는 것을 확인하였으며, 제안하는 방식이 기존 방식에 비해 음성의 존재 여부에 대한 오검출을 줄일 수 있음을 검증하였다.

감사의 글

본 연구는 서울시 산학연 협력사업(WR080951)의 연구결과로 수행되었습니다.

References

1. L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans Antennas Propag, 30, 27-34 (1982).
2. A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," IEEE Trans Audio Speech Lang Processing, **19**, 206-219 (2011).
3. S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans Signal Processing, **49**, 1614-1626 (2001).
4. O. Shalvi and E. Weinstein, "System identification using nonstationary signals," IEEE Trans Signal Processing, **44**, 2055-2063 (1996).
5. I. Cohen, "Relative transfer function identification using speech signals," IEEE Trans Speech Audio Process, **12**, 451-459 (2004).
6. R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," IEEE Trans Audio Speech Lang Processing, **17**, 546-555 (2009).
7. I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," Signal processing, **81**, 2403-2418 (2001).
8. D. Wang and G. J. Brown, *Computational auditory scene analysis: Principles, algorithms, and applications* (Wiley-IEEE Press, New York, 2006), pp. 81-114.
9. G. Hu and D. Wang, "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans Neural Netw, **15**, 1135-1150 (2004).
10. "DARPA Resource Management Continuous Speech Database (RM1)," NIST Speech Disc 2-5.1 (1996).

저자 약력

▸ 신 민 규(Minkyu Shin)

2012년 2월: 고려대학교 전기전자전파 공학부 (학사)
2012년 3월 ~ 현재: 고려대학교 전기전자전파공학부 석/박사 통합

▸ 고 한 석 (Hanseok Ko)

1982년 5월: 미국 카네기 멜론 대학교 전기공학 (공학사)
1986년 5월: 미국 메릴랜드 대학교 시스템 공학(공학석사)
1988년 5월: 미국 존스 홉킨스 대학교 전기공학 (공학석사)
1992년 5월: 미국 카톨릭 대학교 전기공학 (공학박사)
1995년 3월 ~ 현재: 고려대학교 전기전자전파공학부 교수