한국음향학회지 제40권 제5호 pp. 503~514 (2021)
The Journal of the Acoustical Society of Korea Vol.40, No.5 (2021)
https://doi.org/10.7776/ASK.2021.40.5.503

pISSN : 1225-4428
eISSN : 2287-3775

공동 행렬대각화 조건 기반 온라인 음원 신호 분리 및 잔향제거

Online blind source separation and dereverberation of speech based on a joint diagonalizability constraint

유호건,[1] 김도희,[1] 송민환,[2] 박형민[1†]

(Ho-Gun Yu,[1] Do-Hui Kim,[1] Min-Hwan Song,[2] and Hyung-Min Park[1†])

[1]서강대학교 전자공학과, [2]한국전자기술연구원 자율지능IoT연구센터

(Received July 20, 2021; revised September 7, 2021; accepted September 14, 2021)

초 록: 신호에서의 잔향은 암묵음원분리 시스템의 성능을 크게 저하시키는 경향이 있다. 특히 온라인으로 진행되는 시스템일 때, 그 영향이 더욱 두드러진다. 최근 공동 행렬대각화를 활용하여 해당 문제를 해결하고자 하는 연구들이 이루어지고 있다. 본 논문에서는 이를 활용, 발전하여 잔향이 존재하는 환경에서의 미결정 다중 화자의 음원 분리 온라인 알고리즘에 잔향 제거 기능을 추가함으로써 분리한 음원의 품질을 개선하였다. WSJCAM0 데이터베이스에서 실험을 통해 기존에 사용되고 있는 온라인 알고리즘 성능과 비교하였다. 성능 평가는 신호 대 왜곡 비(Signal-to-Distortion Ratio, SDR)와 Perceptual Evaluation of Speech Quality(PESQ)를 통해 이루어졌고, 기존 알고리즘 대비 SDR은 평균 1.23 dB에서 3.76 dB로 향상되었고, PESQ는 1.15에서 2.12로 성능이 향상되었음을 검증하였다.
핵심용어: 온라인 암묵음원분리, 온라인 잔향제거, 독립 성분 분석, 공동 행렬대각화 조건

ABSTRACT: Reverberation in speech signals tends to significantly degrade the performance of the Blind Source Separation (BSS) system. Especially in online systems, the performance degradation becomes severe. Methods based on joint diagonalizability constraints have been recently developed to tackle the problem. To improve the quality of separated speech, in this paper, we add the proposed de-reverberation method to the online BSS algorithm based on the constraints in reverberant environments. Through experiments on the WSJCAM0 corpus, the proposed method was compared with the existing online BSS algorithm. The performance evaluation by the Signal-to-Distortion Ratio and the Perceptual Evaluation of Speech Quality demonstrated that SDR improved from 1.23 dB to 3.76 dB and PESQ improved from 1.15 to 2.12 on average.
Keywords: Online blind source separation, Online dereverberation, Independent component analysis, Joint diagonalizability constraints

PACS numbers: 43.72.Ar, 43.72.Dv

I. 서 론

암묵음원분리(Blind Source Separation, BSS)란 음원의 혼합과정에 대한 사전 정보 없이 동시 다발적으로 발생, 혼합된 음원 신호를 분리하는 것이다. 주파수 영역에서 혼합된 신호를 분리하는 대표적인 기술들로는 주파수영역 독립성분분석(Frequency-Domain Independent Component Analysis, FDICA),[1] 독립벡터분석(Independent Vector Analysis, IVA),[2] 보조함수를 통해 안정성과 필터의 빠른 학습을 적용한 독립벡터분석(auxiliary-function-based IVA, AuxIVA),[3,4] 음원 신호의 분산에 대하여 비음수행렬분해(Nonnega-

†**Corresponding author:** Hyung-Min Park (hpark@sogang.ac.kr)
Department of Electronic Engineering, Sogang University, 35, Baekbeom-ro, Mapo-gu, Seoul 04107, Republic of Korea
(Tel: 82-2-711-8916)

tive Matrix Factorization, NMF)를 적용한 Independent Lowrank Matrix Analysis(ILRMA)[5]들이 있다. 이런 전통적인 암묵음원분리 방법은 국소푸리에변환(Short-Time Fourier Transform, STFT)의 프레임 길이가 잔향 시간보다 충분히 긴 경우에만 성능 저하가 없다는 단점이 있다. 이러한 단점은 weighted prediction error[6,7]와 같은 잔향제거 방법을 통해 프레임 길이보다 긴 잔향 성분을 제거하여 해결할 수 있다. 특히, 최근에는 오프라인 과정에서 암묵음원분리와 잔향제거를 함께 진행하는 방법[8]도 연구되고 있다. 또한, 공분산 행렬에 대하여 공동 행렬대각화 조건을 적용하여 음원 신호의 인접 채널, 주파수, 프레임에 대한 상관도를 고려한 암묵음원분리[9,10]와 공분산 행렬을 full-rank로 추정하는 암묵음원분리[11]에 대한 연구가 있다. 하지만 실제 상황에 대하여 고려한다면 화자가 발화하는 중에 움직이는 상황 뿐 만 아니라, 보청기와 같은 장비는 온라인 동작을 요구한다. 기존 온라인 방식의 암묵음원분리[12-14]와 잔향제거[15]를 적용한 연구가 있다. 본 논문은 오프라인에서 공동 대각화 조건 기반 및 행렬 분해를 통해 암묵음원분리 및 잔향제거 알고리즘 제안과 더 나아가 온라인 방식의 알고리즘을 제안한다.

II. 기존 암묵음원분리 방법

여기서는 전통적인 암묵음원분리로서 rank-1의 공간 모델로 가정한 접근과 잔향을 고려한 이상적인 공간 모델에 대한 접근에 대하여 살펴본다.

2.1 문제 정의

N개의 음원 신호가 혼합된 M개의 다채널 마이크 입력 신호에 대한 국소푸리에변환 영역에서 각각의 시간 프레임 t 와 주파수 인덱스 f 에서의 마이크 입력 신호는

$$\mathbf{x}_{\mathbf{f,t}}=\sum_{\tau=0}^{L_A-1} A_{f,\tau}\mathbf{s}_{f,t-\tau} \tag{1}$$

와 같이 표현된다.[1] 여기서 $\mathbf{x}_{f,t}=[x_{1,f,t},\cdots,x_{M,f,t}]^T$와 $\mathbf{s}_{f,t}=[s_{f,t,1},\cdots,s_{f,t,N}]^T$는 마이크와 음원의 신호에 대한 벡터이며, $[\cdot]^T$는 전치행렬을 의미한다. $A_{f,\tau}\in\mathbb{C}^{MxN}$는 음원에서 마이크까지의 선형시불변 특성을 갖는 전달함수이며, L_A은 해당 필터의 길이를 의미한다. 이 때, N개의 음원 신호를 역으로 추정하기 위한 선형 분리과정[16]은

$$\mathbf{s}_{f,t}=W_{f,0}\mathbf{x}_{f,t}+\sum_{\tau=\Delta}^{\Delta+L-1} W_{f,\tau}\mathbf{x}_{f,t-\tau} \tag{2}$$

와 같이 표현된다.$W_{f,0}$은 분리 행렬이며,$\{W_{f,\tau}\}_{\tau=\Delta}^{\Delta+L-1}$은 잔향제거 행렬이다. $\Delta,L\in\mathbb{N}$은 각각 벽에 의해 반사되어 마이크에 도달하는 초기반사음 시간과 잔향 길이를 나타낸다.

2.2 암묵음원분리에서 rank-1 공간 모델

마이크 입력신호 $\mathbf{x}_{f,t}$는N개의 음원 공간 이미지들 $\{\mathbf{y}_{f,t,n}\}_{n=1}^{N}$의 합으로 구성된다.

$$\mathbf{x}_{f,t}=\sum_{n=1}^{N}\mathbf{y}_{f,t,n}. \tag{3}$$

각각의 음원 신호 $s_{f,t,n}$는 복소정규분포를 따른다고 가정하여

$$s_{f,t,n}\sim N_{\mathbb{C}}(0,\lambda_{f,t,n}) \tag{4}$$

로 표현할 수 있다. $\lambda_{f,t,n}$은 n번째 음원 신호에 대한 파워 스펙트럼의 분산을 나타낸다. 만약 음원 신호가 점 음원이면, 혼합 모델 A_f는 다음과 같이 rank-1인 특성을 갖게 된다. 즉, 음원 공간 이미지 $\mathbf{y}_{f,t,n}$은

$$\mathbf{y}_{f,t,n}=\mathbf{a}_{f,n}s_{f,t,n} \tag{5}$$

으로 표현되며 $\mathbf{a}_{f,n}$은 A_f의 n번째 열벡터에 해당한다. 음원 $\mathbf{y}_{f,t,n}$의 확률 분포는

$$\mathbf{y}_{f,t,n}\sim N_{\mathbb{C}}(\mathbf{0},\lambda_{f,t,n}G_{f,n})\approx N_{\mathbb{C}}(\mathbf{0},R_{f,t,n}) \tag{6}$$

Table 1. Glossary and definition of variables.

#	Term	Definition	dimension
1	N	the number of source signals	-
2	M	the number of sensers	-
3	$\mathbf{s_{f,t}}$	source signal vector	$\mathbb{C}^{N}$
4	$\mathbf{x_{f,t}}$	microphone signal vector	$\mathbb{C}^{M}$
5	$\mathbf{y_{f,t,n}}$	source spatial image vector	$\mathbb{C}^{M}$
6	$A_{f,\tau}$	convolutional transfer function matrix	$\mathbb{C}^{M\times N}$
7	$\mathbf{a}_{f,n}$	column vector of $A_{f,0}$ (steering vector)	$\mathbb{C}^{M}$
8	$W_{f,\tau}$	coefficient matrix	$\mathbb{C}^{N\times M}$
9	$G_{f,n}$	spatial Covariance Matrix	$\mathbb{C}^{M\times M}$
10	$\lambda_{f,t,n}$	power spectral density	$\mathbb{C}$
11	$R_{f,t,n}$	covariance matrix of source signal	$\mathbb{C}^{M\times M}$

로 표현된다. $G_{f,n}=\mathbf{a}_{f,n}\mathbf{a}_{f,n}^{H}$은 rank-1을 갖는 n번째 음원 신호의 공간분산행렬이며, $R_{f,t,n}\in\mathbb{S}_{+}^{M}$은 공분산 행렬, $[\cdot]^{H}$은 켤레 전치행렬이다. Eqs. (3), (6)과 가우시안 분포의 특징을 통해 마이크 입력 신호의 확률 분포는

$$\mathbf{x}_{f,t}\sim N_{\mathbb{C}}(\mathbf{0},\sum_{n=1}^{N}\lambda_{f,t,n}G_{f,n})\approx N_{\mathbb{C}}(\mathbf{0},\sum_{n=1}^{N}R_{f,t,n}) \quad (7)$$

으로 표현된다. Table 1을 통해 주요 변수들에 대한 설명을 정리하였다.

2.3 잔향을 고려한 이상적인 공간 모델에 기반한 암묵음원분리

음원의 공간 이미지에는 마이크 채널간의 상관관계를 갖고 있기 때문에 공간에 대한 정보를 얻을 수 있고, 혼합된 신호로부터 음원을 분리할 수 있다. 하지만 실제 환경에서는 잔향 성분에 의해 공간에 대한 모델이 복잡해진다. 국소푸리에변환을 위한 윈도우의 길이가 음원과 마이크 간에 주파수 응답보다 충분히 길면 잔향에 의한 영향이 줄어들지만 실제 환경에서는 보통 이 가정이 적절하지 않기 때문에 국소푸리에변환을 통해 완벽하게 인접 시간 프레임간의 상관관계를 제거하지 못한다. 본 논문에서는 수식의 단순화를 위해 아래와 같은 표기법을 정의한다.

$$\mathbf{x}_{f}:=[\mathbf{x}_{f,1}^{T},\cdots,\mathbf{x}_{f,T}^{T}]^{T}\in\mathbb{C}^{TM}. \quad (8)$$

$$\mathbf{x}:=[\mathbf{x}_{1}^{T},\cdots,\mathbf{x}_{F}^{T}]^{T}\in\mathbb{C}^{FTM}. \quad (9)$$

$$\mathbf{y}_{n,f}:=\left[\mathbf{y}_{f,1,n}^{T},\cdots,\mathbf{y}_{f,T,n}^{T}\right]^{T}\in\mathbb{C}^{TM}. \quad (10)$$

$$\mathbf{y}_{n}:=\left[\mathbf{y}_{f,n}^{T},\cdots,\mathbf{y}_{f,n}^{T}\right]^{T}\in\mathbb{C}^{FTM}. \quad (11)$$

음원 신호들은 독립적인 특성[2]을 갖기 때문에

$$p(\{\mathbf{y}_{f,t,n}\}_{f,t,n})=\prod_{n=1}^{N}p(\{\mathbf{y}_{f,t,n}\}_{f,t}) \quad (12)$$

의 식을 만족한다. 이 때 각각의 음원 공간 이미지 $\mathbf{z}_{n}$는 평균이 0이고, 공분산 행렬 $R_{n}\in\mathbb{S}_{+}^{FTM}$을 갖는다. 변량 복소정규분포를 따른다고 가정하면

$$\mathbf{y}_{n}\sim N_{\mathbb{C}}(\mathbf{0},R_{n}) \quad (13)$$

와 같이 표현된다. 이때, $\mathbb{S}_{+}^{K}$는 $K\times K$ 크기를 갖는 에르미트 양의 준정부호행렬이다. Eqs (3), (12)~(13)과 정규분포의 특성을 통해

$$\mathbf{x}\sim N_{\mathbb{C}}(\mathbf{0},\sum_{n=1}^{N}R_{n}) \quad (14)$$

을 갖는다.

결론적으로, $\{R_{n}\}_{n=1}^{N}$이 추정된다면, 음원의 공간 이미지는 다채널 Wiener 필터를 통해

$$\mathbf{y}_{n}=R_{n}(\sum_{n=1}^{N}R_{n})^{-1}\mathbf{x} \quad (15)$$

와 같이 각각의 음원 공간이미지를 추정한다. 하지만 공분산 행렬의 차원은 $N(FTM)^{2}$으로 상당히 많은 수의 매개변수를 최적화하는 문제점이 존재한다.

2.4 공분산 행렬의 공동 행렬대각화 조건

공분산 행렬의 차원을 줄이기 위하여 N개의 공분산 행렬 $\{\mathrm{R_n}\}_{\mathrm{n=1}}^{\mathrm{N}}$을 공동으로 대각화하는 방법[9-11]을 적용하여 나타내면

$$\mathrm{P}^{\mathrm{H}}\mathrm{R_n}\mathrm{P}=\mathrm{diag}(\boldsymbol{\lambda_n}) \tag{16}$$

와 같다. 이때 $\mathrm{P}\in\mathbb{C}^{\mathrm{FTM}\times\mathrm{FTM}}$은 정칙행렬이며, $\boldsymbol{\lambda}_{\mathrm{n}}\in\mathbb{R}_{+}^{\mathrm{FTM}}$은 비음수 벡터이다. Eqs. (14)과 (16)로부터

$$\begin{aligned}\mathrm{P}^{\mathrm{H}}\mathbf{x}&\sim N_{\mathbb{C}}\left(\mathbf{0},\sum_{n=1}^{N}\mathrm{P}^{\mathrm{H}}\mathrm{R_n}\mathrm{P}\right)\\&\sim N_{\mathbb{C}}\left(\mathbf{0},\sum_{n=1}^{\mathrm{N}}\mathrm{diag}(\boldsymbol{\lambda}_n)\right)\end{aligned} \tag{17}$$

이고, 공분산에 대한 비대각성분들이 0이 되어 $\mathrm{P}^{\mathrm{H}}\mathbf{x}$의 요소들이 상관관계가 없는 독립적특성을 갖는다. 따라서 $\mathrm{P}^{\mathrm{H}}\mathbf{x}$를 각각의 음원 신호로 간주할 수 있으며, 공동 행렬대각화 방법으로 인해 $\{\mathrm{R_n}\}_{\mathrm{n=1}}^{\mathrm{N}}$의 매개변수 수는 $\mathrm{N(FTM)^2}$개에서 $\mathrm{(FTM)^2+FTM}$으로 줄어들게 된다. P와 $\boldsymbol{\lambda}_{\mathrm{n}}$의 추정을 위한 마이크 입력 신호의 스펙트럼에 대한 음의 우도비용 함수는 다음과 같다.

$$\begin{aligned}J_{cost}&=-\log\ p(\mathbf{x}|\mathrm{P},\boldsymbol{\lambda}_{\mathrm{n}})\\&=-\log\ p(\mathrm{P}^{\mathrm{H}}\mathbf{x})-\log\ \left|\det\mathrm{P}\mathrm{P}^{\mathrm{H}}\right|.\end{aligned} \tag{18}$$

Eq. (18)의 비용함수가 최소가 될 때, P와 $\boldsymbol{\lambda}_{\mathrm{n}}$을 추정해 공분산 행렬을 구할 수 있다.

2.5 인접 채널 및 인접 시간에 대한 역상관화 (decorrelation)를 이용한 암묵음원분리 및 잔향제거

주파수영역 독립성분분석,[1] 독립벡터분석,[2-4] ILRMA[5]는 마이크 입력 신호와 음원 신호의 수가 같은 상황에서 잘 작동하는 대표적인 암묵음원분리 방법이다. 또한 마이크 입력 신호의 잔향성분을 제거하기 위한 여러 효과적인 잔향제거 알고리즘 기술들도 존재한다.[6,7] 음원 스펙트럼의 인접 채널 및 인접 시간에 대한 역상관 모듈 통합 방법[9-11]으로 잔향을 제거함과 동시에 음원 분리를 수행할 수 있다. 인접 채널 및 시간프레임을 고려하여 식(16)의 정칙행렬 P를 각 주파수에 대하여 T^2개의 $\mathrm{M}\times\mathrm{M}$의 차원을 갖는 블록으로 구성된 블록 상 Toeplitz 행렬 $\{\mathrm{P_f}\}_{\mathrm{f=1}}^{\mathrm{F}}\in\mathbb{C}^{\mathrm{TM}\times\mathrm{TM}}$ 로 정의하고, 행렬의 (α,β)번째 블록은

$$\begin{array}{ll}\mathrm{P}_{\mathrm{f},0}\in\mathbb{C}^{\mathrm{M}\times\mathrm{M}} & (\mathrm{if}\,\alpha-\beta=0)\\ \mathrm{P}_{\mathrm{f},\beta-\alpha-\Delta} & (\mathrm{if}\,\beta-\alpha-\Delta+1\in[1,\cdots,L])\\ \mathrm{O}_{\mathrm{M}\times\mathrm{M}} & (\mathrm{otherwise})\end{array} \tag{19}$$

와 같이 정의한다. 이때 $\mathrm{O}_{\mathrm{M}\times\mathrm{M}}$ 은 $\mathrm{M}\times\mathrm{M}$의 영행렬이다. 따라서 정칙행렬 P는 아래 Eq. (20)과 같이 표현된다.

$$\mathrm{P}=\oplus_{\mathrm{f=1}}^{\mathrm{F}}\mathrm{P_f}=\mathrm{diag}\{\mathrm{P}_1,\cdots,\mathrm{P_F}\}. \tag{20}$$

이때, $\oplus_{\mathrm{f=1}}^{\mathrm{F}}\mathrm{P_f}$은 행렬 $\{\mathrm{P_f}\}_{\mathrm{f=1}}^{\mathrm{F}}$의 블록 대각행렬이다. Eq. (20)을 통해 Eq. (16)은

$$\oplus_{\mathrm{f=1}}^{\mathrm{F}}\mathrm{P}_{\mathrm{f}}^{\mathrm{H}}\mathbf{x}_{\mathrm{f}}\sim N_{\mathbb{C}}\left(\mathbf{0},\sum_{n=1}^{N}diag(\boldsymbol{\lambda}_n)\right) \tag{21}$$

와 같이 표현된다. Eqs. (18)과 (21)로 대각화기 $\mathrm{P_f}$를 최적화하는 비용함수는 다음과 같다.[9]

$$\begin{aligned}J=&\frac{1}{2}\sum_{\mathrm{f,t,m}}^{\mathrm{F,T,M}}\left[\frac{\left|\mathbf{e}_{\mathrm{m}}^{\mathrm{T}}\hat{\mathrm{P}}_{\mathrm{f}}^{\mathrm{H}}\hat{\mathbf{x}}_{\mathrm{f,t}}\right|^2}{\lambda_{\mathrm{f,t,m}}}+\log\lambda_{\mathrm{f,t,m}}\right]\\&-\mathrm{T}\sum_{\mathrm{f=1}}^{\mathrm{F}}\log\left|\det\mathrm{P}_{\mathrm{f},0}\right|.\end{aligned} \tag{22}$$

공동 대각화 $\hat{\mathrm{P}}_f$은 $[\mathrm{P}_{\mathrm{f},0}^{\mathrm{T}},\cdots,\mathrm{P}_{\mathrm{f,L}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{C}^{(\mathrm{L}+1)\mathrm{M}\times\mathrm{M}}$, $\hat{\mathbf{x}}_{\mathrm{f,t}}$는 $[\mathbf{x}_{\mathrm{f,t}}^{\mathrm{T}},\mathbf{x}_{\mathrm{f,t-\Delta}}^{\mathrm{T}},\cdots,\mathbf{x}_{\mathrm{f,t-\Delta-L+1}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{C}^{(\mathrm{L}+1)\mathrm{M}}$ 이며, $\mathbf{e}_{\mathrm{m}}$은 m 번째 항이 1인 단위벡터이다.

III. 제안 방법

공동 행렬대각화 조건을 사용한 기존 방법에서는 인접 채널 및 인접 시간의 상관도를 없애는 하나의 필터 $\hat{\mathrm{P}}_{\mathrm{f}}$를 제안하였다. 하지만 매 시간 프레임마다

필터를 추정하기에는 필터의 차원이 다소 크기 때문에 암묵음원분리 및 잔향제거 된 신호를 추정하는 것이 불안정하다. 따라서 하나의 필터를 추정하는 것보다 행렬분해를 적용하여 잔향제거와 음원분리의 필터로 분해하는 방법을 제안하고 온라인 알고리즘 구현을 제안한다.

3.1 대각화 행렬분해

식(22)의 공동 대각화 $\hat{\mathrm{P}}_f$ 행렬을 $\mathrm{P}_{f,0}\in\mathbb{C}^{\mathrm{M}\times\mathrm{M}}$와 $\overline{\mathrm{P}}_{\mathrm{f}}=[\mathrm{P}_{\mathrm{f,1}}^{\mathrm{T}},\cdots\mathrm{P}_{\mathrm{f,L}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{C}^{\mathrm{LMxM}}$로 분리하여 표현하면,

$$\hat{\mathrm{P}}_{\mathrm{f}}=\begin{bmatrix}\mathrm{P}_{\mathrm{f,0}}\\ \overline{\mathrm{P}}_{\mathrm{f}}\end{bmatrix}=\begin{bmatrix}\mathbf{p_{f,0,1}}\cdots\mathbf{p_{f,0,M}}\\ \overline{\mathbf{p}}_{\mathbf{f,1}}\cdots\overline{\mathbf{p}}_{\mathbf{f,M}}\end{bmatrix} \tag{23}$$

와 같이 표현되며, $\mathbf{p_{f,0,m}},\overline{\mathbf{p}}_{\mathbf{f,m}}$은 각각 $\mathrm{P}_{\mathrm{f,0}}$와 $\overline{\mathrm{P}}_{\mathrm{f}}$의 m번째의 열벡터이다. 공동 대각화 행렬에 대한 구조는 Fig. 1(a)와 같다.

m번째 열벡터에 대하여 행렬분해를 진행하면

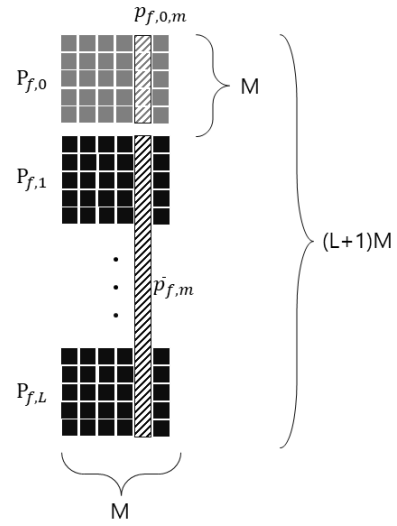
$$\begin{bmatrix}\mathbf{p}_{\mathrm{f,0,m}}\\ \overline{\mathbf{p}}_{\mathrm{f,m}}\end{bmatrix}=\begin{bmatrix}\mathrm{I}_{\mathrm{MxM}}\\ -\overline{\mathrm{L}}_{\mathrm{f,m}}\end{bmatrix}\mathbf{w}_{\mathrm{f,m}} \tag{24}$$

로 $\overline{\mathrm{L}}_{f,m}\in\mathbb{C}^{\mathrm{LMxM}},\mathbf{w}_{\mathrm{f,m}}\in\mathbb{C}^{\mathrm{M}}$, $\mathrm{I}_{\mathrm{MxM}}\in\mathbb{R}^{\mathrm{MxM}}$은 단위행렬로 표현된다. 이 때, Eq. (23)를 통해 아래와 같이 표현된다.

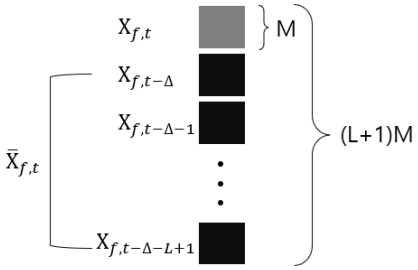
$$\mathbf{z}_{\mathrm{f,t,m}}=\mathbf{x}_{\mathrm{f,t}}-\overline{\mathrm{L}}_{\mathrm{f,m}}^{\mathrm{H}}\overline{\mathbf{x}}_{\mathrm{f,t}}. \tag{25}$$

$$\begin{aligned}\mathbf{e}_{\mathrm{m}}^{\mathrm{T}}\hat{\mathrm{P}}_{\mathrm{f}}^{\mathrm{H}}\hat{\mathbf{x}}_{\mathrm{f,t}}&=\begin{bmatrix}\mathbf{p}_{\mathrm{f,0,m}}\\ \overline{\mathbf{p}}_{\mathbf{f,m}}\end{bmatrix}^{\mathrm{H}}\hat{\mathbf{x}}_{\mathrm{f,t}}\\&=\mathbf{w}_{\mathrm{f,m}}^{\mathrm{H}}(\mathbf{x}_{\mathrm{f,t}}-\overline{\mathrm{L}}_{\mathrm{f,m}}^{\mathrm{H}}\overline{\mathbf{x}}_{\mathrm{f,t}})\\&=\mathbf{w}_{\mathrm{f,m}}^{\mathrm{H}}\mathbf{z}_{\mathrm{f,t,m}}.\end{aligned} \tag{26}$$

$\overline{\mathrm{L}}_{f,m}\in\mathbb{C}^{\mathrm{LMxM}}$의 필터는 단일 음원신호에 대한 잔향제거 필터다. 각 마이크 채널별 잔향 제거된 출력은 $\mathbf{z}_{\mathrm{f,t,m}}$,$\overline{\mathbf{x}}_{\mathrm{f,t}}=[\mathbf{x}_{\mathrm{f,t-\Delta}}^{\mathrm{T}},\cdots,\mathbf{x}_{\mathrm{f,t-\Delta-L+1}}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{C}^{\mathrm{LM}}$은 이전 시간 프레임에 대한 입력신호이며 자세한 구조는 Fig. 1(b)와 같다. $\mathbf{w}_{\mathrm{f,m}}\in\mathbb{C}^{\mathrm{M}}$는 분리행렬 W_{f}의 m번째



(a) diagonalizer matirx $\hat{\mathrm{P}}_f$ structure



(b) ovserved mixture $\hat{\mathbf{x}}_{f,t}$ structure

Fig. 1. Diagonalizer matix $\hat{\mathrm{P}}_f$ and observed mixture $\hat{\mathbf{x}}_{f,t}$ structure.

열벡터이다. Eqs. (22)와 (26)을 통해 최적화 함수는

$$J=\frac{1}{2}\sum_{f,t,m}^{F,T,M}\left[\frac{\left|\mathbf{w}_{\mathrm{f,m}}^{\mathrm{H}}\mathbf{z}_{\mathrm{f,t}}\right|^2}{\lambda_{f,t,m}}+\log\lambda_{f,t,m}\right]-T\sum_{f=1}^{F}\log\left|\det\mathrm{W}_{\mathrm{f}}\right| \tag{27}$$

과 같다. 음원 분리행렬 W_{f}를 업데이트 하는 수식은 보조함수를 이용한 기존 방법[3,4]과 같다. 이 방식은 기존의 경사하강법의 방식[17]보다 안정적이고 빠르게 수렴한다.

$$\mathbf{w}_{\mathrm{f,m}}\leftarrow(\mathrm{W}_{\mathrm{f}}\mathrm{V}_{\mathrm{f,m}})^{-1}\mathrm{e}_{\mathrm{m}}. \tag{28}$$

$$\mathbf{w}_{f,m} \leftarrow \frac{\mathbf{w}_{f,m)}}{\sqrt{\mathbf{w}_{f,m}^{H} V_{f,m} \mathbf{w}_{f,m}}}, \tag{29}$$

여기서 $V_{f,m}$은 아래 Eq. (30)이다.

$$V_{f,m} = \frac{1}{T}\sum_{t=1}^{T} \frac{\mathbf{z}_{f,t,m}\mathbf{z}_{f,t,m}^{H}}{\lambda_{f,t,m}} \in \mathbb{S}_{+}^{M}. \tag{30}$$

선형 예측 필터 $\bar{L}_{f,m}$를 업데이트하는 수식은 Eq. (27)를 $\bar{L}_{f,m}$로 편미분하여 구할 수 있다.

$$\begin{aligned} &\frac{\partial J}{\partial \bar{L}_{f,m}} \\ &= \frac{1}{2}\mathbf{w}_{f,m}^{H}\left(\begin{array}{r} \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\bar{\mathbf{x}}_{f,t}\bar{\mathbf{x}}_{f,t}^{H}}{\lambda_{f,t,m}}\right]\bar{L}_{f,m} \\ -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\bar{\mathbf{x}}_{f,t}\mathbf{x}_{f,t}^{H}}{\lambda_{f,t,m}}\right] \end{array} \right)\mathbf{w}_{f,m} \\ &= 0. \end{aligned} \tag{31}$$

Eq. (31)를 통해 선형 예측 필터 $\bar{L}_{f,m}$는 다음과 같다.

$$K_{f,m}^{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\bar{\mathbf{x}}_{f,t}\bar{\mathbf{x}}_{f,t}^{H}}{\lambda_{f,t,m}} \in \mathbb{C}^{LM\times LM}. \tag{32}$$

$$k_{f,m}^{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\bar{\mathbf{x}}_{f,t}\mathbf{x}_{f,t}^{H}}{\lambda_{f,t,m}} \in \mathbb{C}^{LM\times M}. \tag{33}$$

$$\bar{L}_{f,m} = K_{m}^{\mathbf{x}^{-1}} k_{m}^{\mathbf{x}} \in \mathbb{C}^{LM} \times M. \tag{34}$$

3.2 온라인에서의 최적화

앞서 설명한 오프라인의 방식인 batch processing 알고리즘은 프레임 전반에 걸쳐$(t=1,\cdots,T)$ 얻어진 입력 신호를 통해 필터를 추정한다. 하지만 이러한 시스템은 실제 환경에서와 같이 화자의 위치가 고정되지 않고 발화하는 비정상 음원에 대해서는 채널 간 및 프레임 간의 상관관계가 변하기 때문에 잔향 제거 및 암묵음원분리 성능이 저하된다. 또한 보청기와 같은 음원향상 장치에서는 온라인 동작을 요구한다는 점이다. 이러한 점을 고려하여 앞서 제안한 오프라인 방식 대신에 매 프레임마다 필터를 업데이트하며 분리된 음원을 출력하는 온라인 방식의 알고리즘을 제안한다.

온라인 암묵음원분리를 위해 재귀최소자승법(Recursive Least Squares, RLS)[12-14]을 사용하여, 현재 시간 프레임t의 $V_{f,t,m}$을 이전 시간 프레임의 $V_{f,t-1,m}$을 통해 재귀적으로 계산한다. 따라서 Eq. (30)의 $V_{f,t,m}$는

$$V_{f,t,m} = \alpha V_{f,t-1,m} + (1-\alpha)\frac{\mathbf{z}_{f,t}\mathbf{z}_{f,t}^{H}}{\lambda_{f,t,m}} \tag{35}$$

와 같이 계산되고, α $(1 \le \alpha < 1)$는 망각인자로 과거 신호에 대한 비중을 조절하는 요소이다. 또한, Eq. (28)의 역행렬 연산은 연산비용이 크기 때문에 실시간 동작에서 적합하지 않다. 이를 해결하기 위해 아래 식의 matrix inversion lemma[18]를 이용한다.

$$(B+CD)^{-1} = B^{-1} - B^{-1}C(I+DB^{-1}C)^{-1}DB^{-1}. \tag{36}$$

이 때,

$$(W_{f,t}V_{f,t,m})^{-1} = V_{f,t,m}^{-1}W_{f,t}^{-1} = U_{f,t,m}A_{f,t} \tag{37}$$

와 같이 역행렬 행렬$U_{f,t,m}, A_{f,t}$을 설정하면 Eq. (36)를 사용하여 유도하면 각각의 역행렬들은 이전 시간 프레임에 대해

$$U_{f,t,m} = \frac{1}{\alpha}U_{f,t-1,m} - \frac{p_{f,t}U_{f,t-1,m}\mathbf{x}_{f,t}\mathbf{x}_{f,t}^{H}U_{f,t-1,m}^{H}}{\alpha^{2}+\alpha p_{f,t}\mathbf{x}_{f,t}^{H}U_{f,t-1,m}\mathbf{x}_{f,t}}. \tag{38}$$

$$A_{f,t} \leftarrow A_{f,t} - \frac{A_{f,t}\mathbf{e}_{m}\triangle\mathbf{w}_{f,t,m}^{H}A_{f,t}}{1+\triangle\mathbf{w}_{f,t,m}^{H}A_{f,t}\mathbf{e}_{m}} \tag{39}$$

와 같이 매 프레임마다 추정된다. $\triangle\mathbf{w}_{f,t,m}$은 $W_{f,t}$의 m번째 열벡터 $\mathbf{w}_{f,t,m} \in \mathbb{C}^{M}$의 업데이트 전과 후의 차이를 나타내며 아래와 같이 반영된다.

$$W_{f,t} \leftarrow W_{f,t} + \mathbf{e}_m \triangle \mathbf{w}_{f,t,m}^{H}. \quad (40)$$

다음으로 온라인 잔향제거[15]의 경우에는 이전과 같은 방식으로 Eq. (32)의 $LM \times LM$의 차원을 갖는 $K_{f,m}^{\mathbf{x}}$ 의 역행렬 연산이 음원 분리보다 더 큰 연산비용을 갖게 된다. 마찬가지로 재귀최소자승법의 방식을 적용하여 다음과 같이 나타낼 수 있다.

$$K_{f,t,m}^{\mathbf{x}} = \beta K_{f,t-1,m}^{\mathbf{x}} + \frac{\overline{\mathbf{x}}_{f,t}\overline{\mathbf{x}}_{f,t}^{H}}{\lambda_{f,t,m}}. \quad (41)$$

$$k_{f,t,m}^{\mathbf{x}} = \beta k_{f,t,m}^{\mathbf{x}} + \frac{\overline{\mathbf{x}}_{f,t}\mathbf{x}_{f,t}^{H}}{\lambda_{f,t,m}}. \quad (42)$$

또한, matrix inversion lemma를 통해 $K_{f,t,m}^{\mathbf{x}^{-1}}$ 을

$$Q_{f,t} \leftarrow \frac{K_{f,t-1,m}\overline{\mathbf{x}}_{f,t}}{\beta\lambda_{f,t,m} + \overline{\mathbf{x}}_{f,t}^{H}K_{f,t-1,m}^{\mathbf{x}^{-1}}\overline{\mathbf{x}}_{f,t}}. \quad (43)$$

$$K_{f,t,m}^{\mathbf{x}^{-1}} \leftarrow \frac{K_{f,t-1,m}^{\mathbf{x}^{-1}} - Q_{f,t}\overline{\mathbf{x}}_{f,t}^{H}K_{f,t-1,m}^{\mathbf{x}^{-1}}}{\beta}. \quad (44)$$

$$\overline{L}_{f,t,m} = \overline{L}_{f,t-1,m} + Q_{f,t}\mathbf{z}_{f,t,m}^{H} \quad (45)$$

와 같이 매 프레임마다 추정할 수 있다. 온라인 알고리즘에서 잔향제거 부분에서의 $\lambda_{f,t,m}^{W\text{-}P\text{-}E}$는 전 시간 프레임을 통해 업데이트된 필터를 통해 다음과 같이 추정할 수 있다.

$$\lambda_{f,t,m}^{W\text{-}P\text{-}E} \leftarrow \mathbf{w}_{f,t-1,m}^{H}(\mathbf{x}_{f,t} - \overline{L}_{f,t-1,m}^{H}\overline{\mathbf{x}}_{f,t}). \quad (46)$$

이 때, $\lambda_{f,t,m}^{BSS}$는 음원 신호가 정규분포를 따른다고 가정하여 다음과 같이 계산할 수 있다.

$$\lambda_{f,t,m}^{BSS} = \frac{\sum_{f=1}^{F}\left|\mathbf{w}_{f,t-1,m}^{H}\mathbf{z}_{f,t,m}\right|^2}{F} \quad (47)$$

IV. 실 험

4.1 평가 지표

첫 번째 지표는 신호 대 왜곡 비(Signal-to-Distortion Ratio, SDR)[19]이다. 즉, 마이크에 들어온 입력 신호를 암묵음원분리를 통해 얻은 해당 음원 clean 신호 sig_{target}와 해당 음원 출력신호 sig_{output}의 power 비로 아래의 식과 같다.

$$SDR = 10\log_{10}\frac{\|sig_{target}\|^2}{\|sig_{output} - sig_{target}\|^2}. \quad (48)$$

두 번째 지표는 Perceptual Evaluation of Speech Quality (PESQ)[20]이다. 이 지표는 해당 음원 신호와 암묵음분리를 통한 해당 신호 간의 유사도를 인지적 특성을 반영하여 측정하는 방식이다. PESQ는 주관적 음질 평가 방법을 대체할 수 있는 객관적 음질평가로 만점인 4.5점에 가까울수록 사람들은 음질이 높다고 느낀다.

4.2 실험 환경

본 실험은 WSJCAM0 데이터베이스[21]를 기반으로 음원 신호를 구성했고, 잔향이 존재하는 입력신호는 음원으로부터 마이크 위치까지의 임펄스 응답을 image method[22]에 따라 음원 신호에 합성 곱하여 혼합입력 신호를 생성하였다. 이때, 음원신호와 마이크는 각각 2개, 6개로 구성하고, 혼합하는 음원들은 서로 중복되지 않고, 임의로 선택하였다. 구체적인 실험 환경은 Fig. 2와 같다.

마이크 어레이는 0.04 m 간격으로 일렬로 위치시켰다. 마이크 어레이의 중심은 [2.5 m, 2.5 m, 1 m]에 존재하고, 음원의 거리는 1 m이고, 음원의 각도는 중심선을 기준으로 30°, -80°를 이룬다. 방의 크기는 5 m × 4 m × 3 m이다. 이때, 잔향 시간(RT_{60})은 잔향시간이 작은 0.2 s부터 잔향시간이 큰 1 s로 0.2 s 간격으로 설정했다. 그리고 각 음원과 마이크 위치 사이의 임펄스 응답을 합성 곱하며 잔향 별로 동일한 음원 데이터를 생성하였다. 마이크 입력신호의 샘플링 주파수는 16 kHz이며, 국소푸리에변환에서 Hanning 윈도

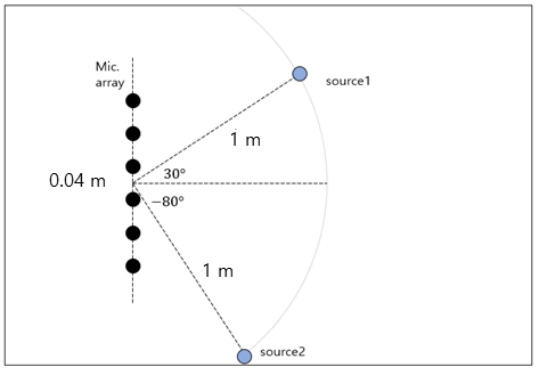


Fig. 2. (Color available online) Recording conditions of impulse response obtained from image method.

우 및 윈도우 프레임 길이와 프레임 간 간격은 각각 64 ms, 16 ms로 설정하였다. 필터 및 매개변수의 초기 값에 대해서는 $W_{f,t=0}$와 $A_{f,t=0}$는 I_{MxM}, $K_{f,t=0,m}^{-1}$는 $10^{-5} \cdot I_{LMxLM}$ $\overline{L}_{f,t=0,m} = O_{LMxM}$ 및 매개변수 $\{\alpha,\beta\}$는 {0.98, 0.99}로 설정하였다.

4.3 실험 결과

본 실험은 기존 온라인 IVA[13]와 제안 알고리즘을 두 가지 지표를 통해 비교하였다. 두 방법 모두 암묵 음원분리에서 사용되는 음원 파워 스펙트럼 밀도 λ^{BSS}의 값은 정규분포를 따른다는 가정으로 동일하게 설정하였다.

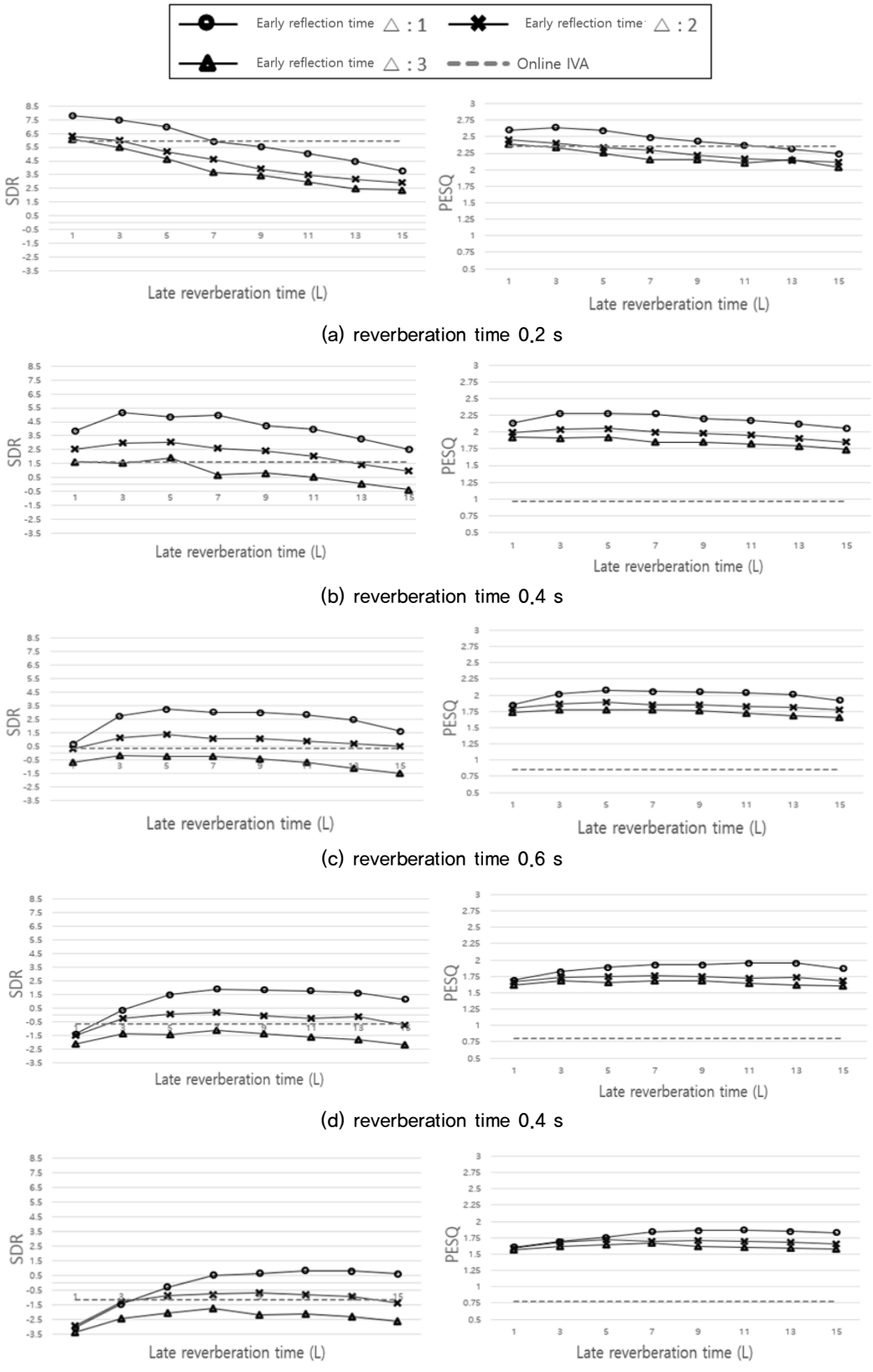
잔향 환경에서 초기 반사는 음성인식에 있어서 사람의 명료도를 향상시키고,[23] 음성인식(ASR) 성능을 향상시킨다.[24] 따라서 초기 반사음 및 잔향 시간 $\{\triangle, L\}$의 값을 고려하여 잔향 시간이 짧은 0.2 s부터 비교적 긴 1 s의 실험 환경에서 SDR과 PESQ의 평균 값 성능을 평가한다.

위의 Table 2은 실험을 통해 가장 높은 성능을 나타내는 초기반사음과 잔향길이를 고려한 실험 결과이다. 기존 온라인 IVA보다 제안한 방법의 성능이 모두 높은 것을 확인 할 수 있다. Fig. 3은 초기 반사음 및 잔향시간에 따른 성능 추이 그래프이다. 첫 번째로 초기 반사음에 따른 성능 추이를 살펴보면, 초기 반사음의 길이를 $\triangle$=1로 설정할 경우 가장 높은 성능을 나타내는 것을 확인할 수 있다. 또한 초기 반사음의 길이가 길수록 분리 성능이 낮아지는 것을 확인할 수 있다. 이러한 점은 초기 반사음을 길게 설정할 경우 반사되어 돌아오는 잔향신호 성분이 남아있기 때문에 성능이 낮아지게 된다. 두 번째로 잔향시간에 따른 성능 추이를 살펴보면, 잔향이 커짐에 따라 최적의 필터 길이가 길어짐을 알 수 있다. 즉 잔향에 영향이 클수록 고려해야 하는 이전 시간의 입력 또한 길어진다는 것이다.

Table 2. Source separation performance in terms of SDR, PESQ according to reverberation time.

Method	Online IVA	Proposed method
	SDR (dB) / PESQ	SDR (dB) / PESQ
0.2 s	5.93 / 2.35	**7.77 / 2.60**
0.4 s	1.60 / 0.97	**5.18 / 2.28**
0.6 s	0.33 / 0.86	**3.13 / 2.06**
0.8 s	-0.67 / 0.80	**1.87 / 1.94**
1 s	-1.19 / 0.78	**0.82 / 1.87**
average	1.12/1.15	**3.75/2.15**

다양한 잔향 실험에서 기존의 온라인 분리 방법보다 SDR과 PESQ 모두 높은 성능을 확인할 수 있다. 하지만 온라인 방식은 시간경과에 따른 재귀적 방식을 사용하기 때문에 잔향의 영향이 큰 환경일수록 잔향의 영향이 적은 환경보다 성능이 낮아지는 것을 확인할 수 있다. 다음 실험은 시간 경과에 따른 SDR 및 PESQ의 성능을 통해 온라인 방식에서의 시간에 따른 각 온라인 분리방법의 암묵음 분리 성능을 살펴보았다. Fig. 4는 Fig. 3의 실험 결과를 통해 각 잔향 환경마다 높은 성능을 나타내는 최적의 초기 반사음과 잔향 시간을 설정하여 실험하였다. Fig. 4의 결과를 살펴보면, 기존의 Online-IVA는 잔향의 영향이 적은 환경(RT_{60} = 0.2 s)에서는 시간에 따른 분리성능이 향상되지만, 잔향의 영향이 커질수록 제대로 분리가 되지 않는 것을 확인 할 수 있다. 기존의 방법과 비교하여, 제안한 방법을 살펴보면 초기 시간에는 분리 성능이 떨어지지만 시간의 경과에 따라서 분리 성능이 점차 향상되는 것을 확인할 수 있다. 상단의 Fig. 5는 잔향 시간이 0.4 s인 실험 환경에서의 음원 분리 결과 스펙트로그램의 예시이다. 온라인 방식으로 인해 두 방법 모두 초기 시간에서는 신호의 분리가 뚜렷하게 나타지 않는다. 하지만 기존 방법(c)에서는 시간이 경과해도 목표 음원 신호에 가깝게 분리되지 않지만 제안 방법(d)에서는 목표 음원신호(b)에 가깝게 분리된 것을 확인 할 수 있다.



(a) reverberation time 0.2 s

(b) reverberation time 0.4 s

(c) reverberation time 0.6 s

(d) reverberation time 0.4 s

(e) reverberation time 1 s

Fig. 3. Online source separation performance according to late-reverberation and early reflection.

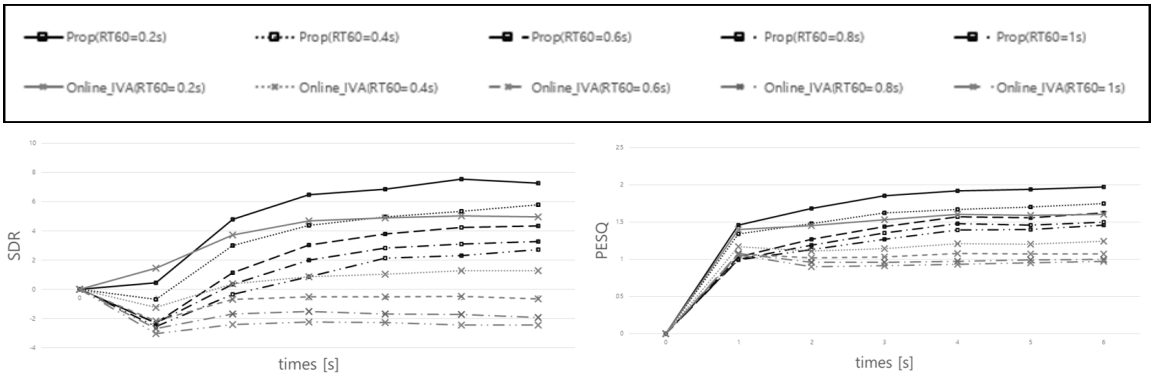
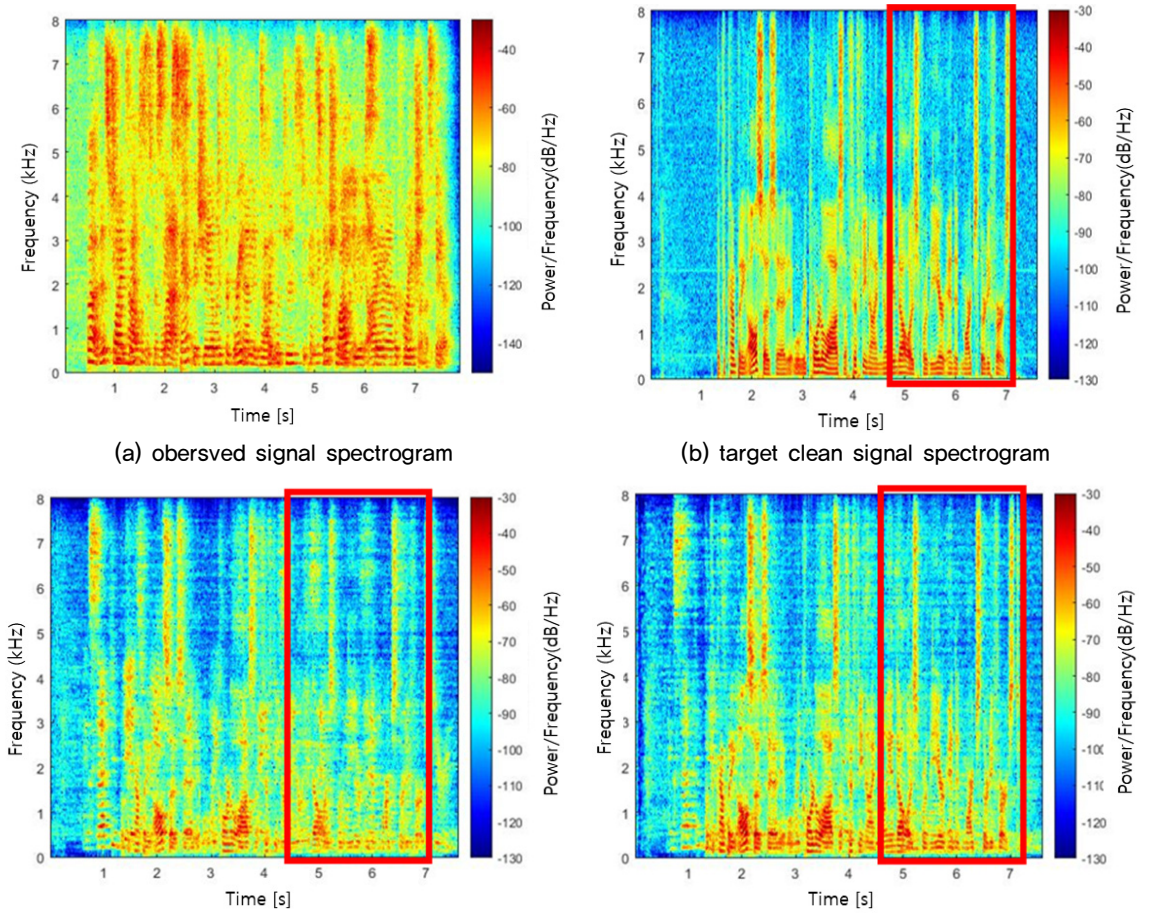


Fig. 4. Online source separation performance over time.



(a) obersved signal spectrogram (b) target clean signal spectrogram

(c) online IVA output signal spectrogram (d) proposed method output siganal spectrogram

Fig. 5. (Color available online) Spectrogram of (a) a reverberant mixture, spectrogram of (b) a clean signal and spectrograms of separated signals obtained by (c) online IVA and (d) proposed method.

V. 결 론

본 연구에서는 공동 행렬대각화의 행렬 분해를 통해 잔향 성분에 대한 상관도를 줄이는 방법을 제안했고, 또한 온라인 암묵음원분리 및 잔향제거 알고리즘을 제안하였다. 실험 결과 제안된 온라인 방식은 잔향이 존재하는 다중화자 발화 환경에서 기존의 암묵음원분리 알고리즘보다 우수한 분리 성능을 보

이는 것을 확인하였다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2020-0-00860, 음향기반 멀티-롤 지원 초소형 재난·안전용 센서 디바이스 및 재난상황 인식 기술개발 및 2019-0-01376, 다중 화자간 대화 음성인식 기술개발).

References

1. P. Smaragdis,“Blind separation of convolve mixtures in the frequency domain,” Neurocomput. **22**, 21-34 (1998).
2. T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher order frequency dependencies,” IEEE Trans. ASLP. **15**, 70-79 (2007).
3. N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. 189-192 (2011).
4. N. Ono and S. Miyabe, “Auxiliary-function-based independent component analysis for super-Gaussian sources,” Proc. Int. Conf. Latent Variable Anal. Signal Separation, 165-172 (2010).
5. D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” IEEE/ACM Trans. ASLP. **24**, 1626-1641 (2016).
6. T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B. H. Juang, “Blind speech dereverberation with multi-channel linear prediction based on short time fourier transform representation,” Proc. ICASSP. 85-88 (2008).
7. T. Yoshioka and T. Nakatani, “Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening,” IEEE Trans. Audio, Speech Lang. Process. **20**, 2707-2720 (2012).
8. T. Nakatani, C. Boeddeker, K. kinoshita, R. Ikeshita, M. Delcroix, and R. Haeb-Umbach, “Jointly optimal denoising, dereverberation, and source separation,” IEEE/ACM Trans. ASLP. **28**, 2276-2282 (2020).
9. R. Ikeshitam N. Ito, Nakatani, and H. Sawada, “A unifying framework for blind source separation based on a joint diagonalizability constraint,” Proc. Eur. Signal Process. Conf. 1-5 (2019).
10. R. Ikeshita, N. Ito, T.Nakatani, and H. Sawada, “Independent low-rank matrix analysis with decorrelation learning,” Proc. IEEE WASPAA. 288-292 (2019).
11. K. Sekiguchi, Y. Bando, A. Nugraha, K. Yoshiim, and T. Kawahara, “Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation,” IEEE/ACM Trans. ASLP. 28, 2610-2625 (2020).
12. M. T. Akhtar, T.-P. Jung, S. Makeig, and G. Cauwenberghs, “Recursive independent component analysis for online blind source separation,” IEEE Int. Symp. Circuits Syst. **6**, 2813-2816 (2012).
13. T. Taniguchi, N. Ono, A. Kawamata, and S. Sagayama, “An auxiliary-function approach to online independent vector analysis for real-time blind source separation,” Proc. HSCMA. 107-111 (2014).
14. S.-H. Hsu, T. Mullen, T.-P. Jung, and G. Cauwenberghs, “Online recursive independent component analysis for real-time source separation of high-density EEG,” Proc. IEEE Eng. Med. Biol. Soc. Conf. 3845-3848 (2014).
15. T. Yoshioka and T. Nakatani, “Dereverberation for reverberation-robust microphone arrays,” Proc. Eur. Signal Process. Conf. 1-5 (2013).
16. T. Nakatani and K. Kinoshita, “A unified convolutional beamformer for simultaneous denoising and dereverberation,” IEEE Signal Processing Letters, **26**, 903-907 (2019).
17. S.-I. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” Adv. Neural Inf. Process. Syst. **8**, 752-763 (1996).
18. M. Woodbury, “Inverting modified matrices,” Memorandum Rep. **42**, MR0038136 (1950).
19. E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source,” IEEE Trans. Audio, Speech, and Lang. Process. **14**, 1462-1469 (2006).
20. A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, “Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs,” Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. **2**, 749-752 (2001).
21. T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, “WSJCAM0: A british english speech corpus for large vocabulary continuous speech recognition,” Proc. ICASSP. 81-84 (1995).
22. J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” J. Acoust. Soc. Am. **65**, 943-950 (1979).
23. S. Bradley, H. Sato, and M. Picard, “On the impor-

tance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233-3244 (2003).
24. T. Nishiura, Y. Hirano, Y. Denda, and M. Nakayama, "Investigations into early and late reflections on distant-talking speech recognition toward suitable reverberation criteria," Proc. Interspeech, 1082-1085 (2007).

저자 약력

▸ 유 호 건 (Ho-Gun Yu)

2018월 2월 : 서강대학교 전자공학과 학사
2020월 9월 ~ 현재 :서강대학교 전자공학과 석사과정

▸ 김 도 희 (Do-Hui Kim)

2021월 2월 : 서강대학교 전자공학과 학사
2021월 3월 ~ 현재 :서강대학교 전자공학과 석사과정

▸ 송 민 환 (Min-Hwan Song)

2003년 2월 :건국대학교 정보통신공학과 학사
2005년 8월 : 건국대학교 정보통신공학과 석사
2005년 11월 ~ 현재 : 한국전자기술연구원 자율지능IoT연구센터 책임연구원

▸ 박 형 민 (Hyung-Min Park)

1997년 2월 : KAIST 전기 및 전자공학과 학사
1999년 2월 : KAIST 전기 및 전자공학과 석사
2003년 8월 : KAIST 전자전산학과 박사
2003년 9월 ~ 2005년 2월 : KAIST 바이오시스템학과 박사 후 연수 과정
2005년 3월 ~ 2007년 1월 : Carnegie Mellon University, Language Technologies Institute 박사 후 연수 과정
2007년 3월 ~ 2011년 2월 : 서강대학교 전자공학과 조교수
2011년 3월 ~ 2016년 2월 : 서강대학교 전자공학과 부교수
2016년 3월 ~ 현재 : 서강대학교 전자공학과 교수