Journal of KIIT. Vol. 16, No. 10, pp. 69-74, Oct. 31, 2018. pISSN 1598-8619, eISSN 2093-7571 69
http://dx.doi.org/10.14801/jkiit.2018.16.10.69

음성신호와 잡음신호의 분리를 위한 독립벡터분석에 기초한 블라인드 음원분리방법

최 재 승*

A Blind Source Separation Method Based on Independent Vector Analysis for Separation of Speech Signal and Noise Signal

Jae-Seung Choi*

요 약

마이크로폰으로부터 입력되는 복수의 음원을 인식하기 위해서는 여러 음원 중에서 특별히 원하는 음원만을 추출하는 음원분리 기술이 필요하다. 본 논문에서는 독립벡터분석에 의한 음원분리 방법을 제안하며, 제안한 방법은 두 채널의 마이크로폰으로 입력된 음성신호와 자동차 잡음을 분리한다. 제안한 독립벡터분석 방법은 시간영역의 신호를 주파수영역의 신호로 바꾸어 각 주파수의 분리행렬을 이용하여 원래의 음성신호와 자동차 잡음신호를 분리한다. 본 실험에서는 제안한 주파수영역의 독립벡터분석 방법을 이용하여 음성 및 잡음의 두 음원에 대하여 파형 및 스펙트로그램의 음원분리 실험을 통하여 향상 효과를 검증하였다. 실험결과로부터 본 논문에서 제안한 독립벡터분석 방법이 혼합 음성신호와 자동차 잡음신호를 상당히 깨끗하게 원래의 신호로 분리한 것을 확인할 수 있었다.

Abstract

In order to recognize a plurality of input sources from microphones, a source separation technique is needed to extract only the desired source among various sources. This paper proposes a source separation method by Independent Vector Analysis (IVA), and it separates a speech signal and car noise inputted from two channel microphones. The proposed IVA method converts a time domain signal into a frequency domain signal, and separates the original speech signal and the car noise signal using a separation matrix in each frequency. In this experiment, the improvement effect was verified by the source separation experiments from waveform and spectrogram for two sources of the speech and noise using the proposed IVA method in the frequency domain. From the experimental results, it was confirmed that the proposed IVA method separates the mixed speech signal and the car noise signal into the original signal in a fairly clean state.

Keywords
source separation, independent vector analysis, independent component analysis, microphone, car noise

* 신라대학교 스마트전기전자공학부 교수
- ORCID: http://orcid.org/0000-0002-5699-9701

· Received: Aug. 09, 2018, Revised: Sep. 06, 2018, Accepted: Sep. 09, 2018
· Corresponding Author: Jae-Seung Choi
Div. of Smart Electrical and Electronic Engineering, Silla University, 140 Baegyang-daero(Blvd), 700beon-gil(Rd), Sasang-gu, Busan, 46958 Korea,
Tel.: +82-51-999-5608, Email: jschoi@silla.ac.kr

Ⅰ. 서 론

최근 컴퓨터에 의한 음성인식 기술은 계속 진보되고 있으며 깨끗한 환경에서 마이크로폰을 향해서 녹음하는 경우에는 상당히 높은 수준의 음성인식 결과를 구하고 있다. 그러나 다른 한편으로는 다양한 배경음, 예를 들면 주변 사람의 음성, 음악, 소음 등이 있는 배경환경에서는 음성인식 성능이 급격히 떨어지는 현상이 자주 나타난다[1-3]. 마이크로폰으로부터 입력되는 복수의 음성을 인식하는 경우에는 목적으로 하는 음성과 방해하는 음성과의 혼합 및 잔향 등의 영향이 문제가 되고 있다. 이러한 상황에서 청취하고자 음성을 인식하기 위해서는 여러 음성 중에서 특별히 원하는 음성만을 분리 추출하는 음원분리 기술이 필요하다. 이 음원분리기술은 다양한 음원이 존재하는 환경에서 음성인식 시스템에 적절한 입력을 주기 위한 중요한 요소 기술이다[4].

근년 마이크로폰으로 입력되는 음성 중에서 목적으로 하는 음성만을 구별하는 음원분리 기술로서는 독립성분 분석(ICA, Independent Component Analysis)에 의한 블라인드 음원분리(BSS, Blind Source Separation) 기술이 각광을 받고 있다. ICA는 음원 및 혼합음원의 정보를 사용하지 않고 관측된 신호만을 사용하여 혼합전의 음원을 추정하여 음원을 분리하는 기술이다. 이 ICA는 복수의 음원이 통계적으로 서로 독립적이라는 가정에 근거하여 분리신호가 서로 독립되게 필터를 설계하여 음원을 분리하는 수법이다. 이러한 ICA에 기초한 분리수법은 희망하는 음원과 잡음 사이의 독립성에 착목하여 음원에 관한 사전정보를 사용하지 않는 블라인드 처리방법을 이용하여 녹음한 혼합신호로부터 목적으로 하는 음원을 분리하여 복원하는 방법이다[5].

최근에는 음향신호에서는 혼합계가 수록하는 녹음실의 잔향에 기인하는 컨볼루션 합이 되기 때문에 단시간 푸리에변환(STFT, Short-time Fourier Transform)로부터 구해진 복소 스펙트럼의 각 주파수 빈(bin)에 대해서 ICA를 적용하는 것으로부터 주파수마다 분리행렬을 추정하는 주파수영역 ICA (FDICA, Frequency-domain ICA)가 제안되었다[6]. 이후에 주파수마다 분리행렬의 추정을 달성하는 수법으로써 독립벡터분석(IVA, Independent Vector Analysis)도 제안되었다. IVA는 ICA를 다변량으로 확장한 이론으로써 각 음원의 주파수성분을 하나로 정리한 주파수벡터의 생성모델을 가정하며, 구대칭적인 성질을 가진 비가우스 다변량분포를 가정함으로써 동일음원의 주파수성분간의 고차상관의 고려가 가능한 방법이다[7][8].

마이크로폰에 의하여 혼합신호를 합성하는 방법으로 ISM(Image Source Method)가 주로 사용된다. ISM은 주어진 환경에서 합성된 공간 임펄스 응답(RIR, Room Impulse Response), 즉 음원과 음향 센서 사이의 전달함수를 생성하기 위해 사용되는 기술이다. 이 RIR을 주어진 음원과 컨벌루션함으로써 음성 데이터의 표본을 취득할 수 있다. 본 논문에서는 제안한 IVA에 의한 혼합신호 분리의 성능을 평가하기 위하여 ISM 툴박스를 사용하여 룸에서의 임펄스 응답신호를 만들고 음성과 컨볼루션함으로써 두 채널의 마이크로폰 혼합신호를 만들었다[9].

본 논문에서는 IVA 방법에 의한 음원분리 방법을 제안하며, 두 채널을 통하여 마이크로폰으로 입력된 혼합 음원신호와 자동차 잡음을 분리한다. 제안한 IVA 방법은 시계열 신호를 주파수영역 신호로 변환하여 각 주파수에 대하여 분리행렬을 구하여 원래의 음성신호와 잡음신호를 분리하는 방법이다. 본 논문에서는 실제 환경을 고려하여 주파수영역의 IVA 방법을 이용하여 음성 및 잡음의 두 음원에 대하여 파형 및 스펙트로그램의 음원분리 실험을 통하여 향상 효과를 검증한다.

Ⅱ. IVA에 의한 음원분리

IVA는 다채널의 주파수영역에 있어서 음원을 분리하는 방법이며 ICA의 확장수법이다. 즉, IVA는 개선된 FDICA이며 각 주파수 빈에 대한 독립성보다는 주파수 빈 사이에 종속성이 존재한다고 간주한다. 그러나 근년 IVA는 독립성을 주파수 빈뿐만 아니라 전대역의 정보를 기준으로 평가하고 있기 때문에 원신호의 주파수 스펙트럼에 고조파와 같은 국소적인 치우침이 없으며, FDICA를 이용하여 독립성을 기준으로 원신호를 분리한다[7][8].

IVA에 의한 음원신호 추정은 FDICA와 상당히 비슷하다고 볼 수 있다. 관측신호 $x_i(t)$는 음원신호

의 혼합행렬에 의하여 식 (1)과 같이 컨볼루션으로 나타낼 수 있다. 여기에서 t는 이산시간, $s_j(t-\tau)$는 τ의 시간지연을 가진 j번째 음원신호, $A_{ij}(\tau)$는 혼합행렬이다.

$$x_i(t)=\sum_{j=1}^{N}\sum_{\tau=0}^{T}A(\tau)_{ij}s_j(t-\tau)=A(t)*s(t) \quad (1)$$
$$\text{for } i=1,\cdots,l$$

각 채널의 관측신호를 STFT를 사용하여 시간 프레임마다 주파수 스펙트럼으로 변환하여 복소 진폭값을 구한다. 어떤 시간 프레임에 있어서 j 번째의 신호원의 신호벡터를 s_j로 하였을 때, N개의 음원신호에 대한 i 번째의 혼합신호 x_i 는 식 (2)와 같다.

$$x_i=\sum_{j=1}^{N}Ai_js_j, \text{for } j=1\sim N \quad (2)$$

여기에서 $Ai=[ai^{(1)},ai^{(2)},.....,ai^{(k)}]$는 대각행렬 Ai의 요소벡터이다. 따라서 혼합신호 x_i의 ω번째의 주파수 빈에 의한 혼합신호 $X_i(k,\omega)$는 식 (3)과 같이 표현할 수 있다.

$$X_i(k,\omega)=\sum_{i=1}^{N}A_{ij}(\omega)\,S_j(k,\omega) \quad (3)$$

여기에서 k은 STFT의 각 프레임을 나타내며 ω는 주파수 빈을 나타낸다. 주파수영역의 분리신호 $Y_j(k,\omega)$ 는 식 (4)와 같은 관계식이다.

$$Y_j(k,\omega)=\sum_{j=1}^{M}W_{ji}(\omega)\,X_i(k,\omega) \quad (4)$$

$Y(k,\omega)$는 주파수영역의 분리된 신호이며, $W_{ji}(k,\omega)$는 ω번째 주파수 빈의 분리된 행렬이다. $Y_j(k,\omega)$는 $Y(k,\omega)$의 j번째 주파수 영역의 요소이다. 식 (4)는 (채널×시간)의 2차원 혼합신호로부터 IVA에 의하여 음원신호에 대한 대역별 복소진폭과 분리행렬을 추정한다. 이 경우에 IVA는 STFT로 도출된 일련의 주파수 스펙트럼을 벡터로 보고 전대역 정보를 반영한 독립성을 기준으로 한다. 따라서 IVA에서는 정확한 분리행렬의 추정이 가능하며 음원신호가 정확하게 복원된다. 식 (3), (4)는 식 (5)와 같이 다시 표현할 수 있다.

$$X(k,\omega)=A(\omega)\,S(k,\omega) \quad (5)$$
$$Y(k,\omega)=W(\omega)\,X(k,\omega)$$

추정된 식 (5)의 주파수영역의 분리된 신호 $Y(k,\omega)$는 분리행렬 $w(\tau)$를 사용하여 식 (6)과 같은 시간영역의 음원신호 $y_i(t)$로 나타낼 수 있다.

$$y_i(t)=\sum_{i=1}^{N}\sum_{\tau=0}^{T-1}W(\tau)_{ji}x_i(t-\tau) \quad (6)$$
$$\text{for } j=1,....,l.$$

본 논문에서는 2×2인 경우를 가정하며, 주파수영역에서의 혼합행렬 $A(\omega)$와 분리행렬 $W(\omega)$를 다음 식과 같이 표현한다.

$$A(\omega)=\begin{bmatrix}a_{11}\,a_{12}\\a_{21}\,a_{22}\end{bmatrix} \quad (7)$$

$$W(\omega)=\begin{bmatrix}w_{11}\,w_{12}\\w_{21}\,w_{22}\end{bmatrix} \quad (8)$$

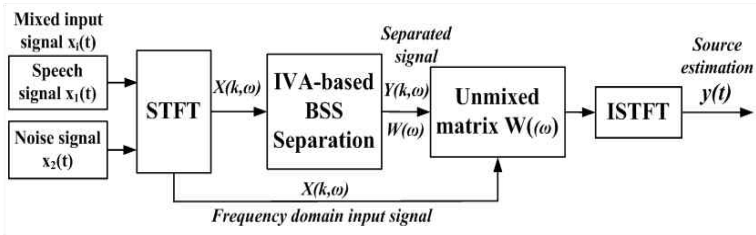


그림 1. 제안한 IVA 음원분리 시스템
Fig. 1. Proposed IVA source separation system

본 논문에서 제안한 IVA 음원분리의 시스템은 그림 1과 같으며, 두 개의 마이크로폰으로 수신되는 음원을 사용하며 각 음원은 서로 결합되어 각 마이크로폰에서 혼합신호를 생성한다. 이 후에 혼합신호에 STFT를 적용하여 시간영역의 혼합신호를 주파수 영역으로 변환한 다음 IVA를 사용하여 음원신호를 분리한다. 이 후에 역 단시간 푸리에변환(ISTFT, Inverse Short-Time Fourier Transform)을 적용하여 주파수 영역의 신호를 시간 영역의 신호로 변환하여 분리된 음원신호를 추정한다.

식 (9)는 2 채널 시간영역의 혼합신호를 STFT 변환을 사용하여 주파수영역의 혼합신호로 변환하는 관계식을 나타낸다.

$$x_i(t) = [x_1(t)\, x_2(t)]^T \rightarrow STFT \rightarrow X(k,\omega) = [X_1(k,\omega)\, X_2(k,\omega)]^T \quad (9)$$

여기에서 $x_i(t)$는 혼합신호를 나타내며 혼합신호 1 ($x_1(t)$)과 혼합신호 2 ($x_2(t)$)로 구성되며, ^{T}는 전치를 나타낸다. 또한 $X(k,\omega)$는 주파수영역의 혼합신호이며, $X_1(k,\omega)$과 $X_2(k,\omega)$로 구성된다.

III. 실험 조건 및 결과

본 논문에서는 ISM 툴박스를 사용하여 음성과 잡음신호를 컨볼루션함으로써 두 채널의 마이크로폰 혼합신호를 만들었다[9]. 본 논문에서 음원으로 사용하는 음성[10]은 표본주파수 16kHz인 남성 3명과 여성 3명의 총 6개의 음성신호를 사용한다. 남성 음성의 길이는 8초에서 11초 사이이며, 여성 음성의 길이는 9초에서 12초 사이이다. 잡음신호는 16kHz의 표본주파수를 가진 Aurora2 데이터베이스[11]의 자동차 잡음(Car Noise)을 사용하였으며, 남성 및 여성 각각 3명인 총 6개의 깨끗한 음성신호에 대해서 SNR이 13.64dB이 되도록 잡음을 중첩시켜 음원신호를 생성하였다. 본 실험에서는 음성과 잡음의 음원의 길이가 서로 다르기 때문에 두 신호 중에서 작은 부분의 음원의 길이에 맞추어 실험을 실시하였다. 모든 실험 신호는 매트랩을 사용하여 생성되었으며, STFT의 분석 프레임 길이는 512 표본이며 384 표본을 중첩시켜 해당 프레임을 이동시킨다.

본 논문에서 제안한 IVA를 사용하여 두 채널에 입력되는 혼합된 음성 및 잡음신호의 두 음원을 분리하는 실험결과를 그림 2~그림 5에 나타낸다. 그림 2는 마이크로폰을 통해 자동차 잡음과 혼합되어 들어온 음성신호 중의 하나인 중첩된 음원신호(그림 2의 상단)와 잡음신호(그림 2의 하단)를 각각 나타내고 있다. 그림 3은 본 논문에서 제안한 IVA에 의하여 분리된 음성(그림 3의 상단) 및 자동차 잡음(그림 3의 하단)을 각각 나타낸다. 그림 4는 자동차 잡음이 중첩되지 않은 깨끗한 음원신호를 나타낸다.

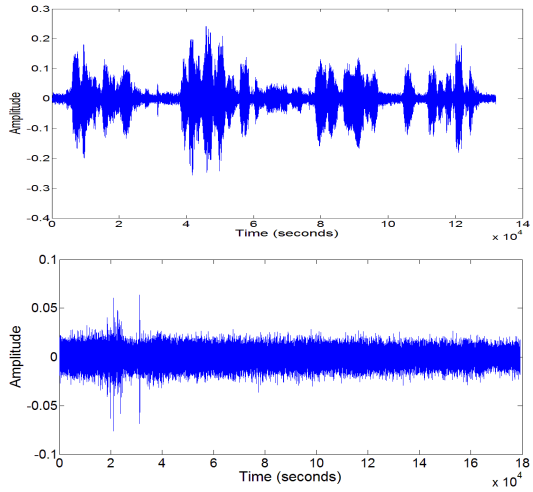


그림 2. 잡음이 중첩된 입력 혼합 음성신호 및 자동차 잡음

Fig. 2. Input mixed noisy speech signal and car noise

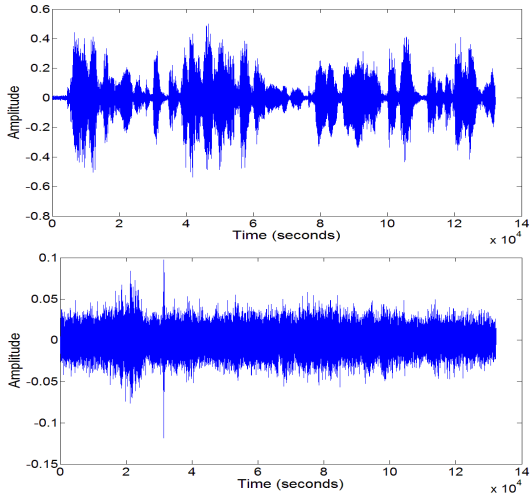


그림 3. IVA에 의해서 분리된 음성신호 및 자동차 잡음

Fig. 3. Separated speech signal and car noise by IVA

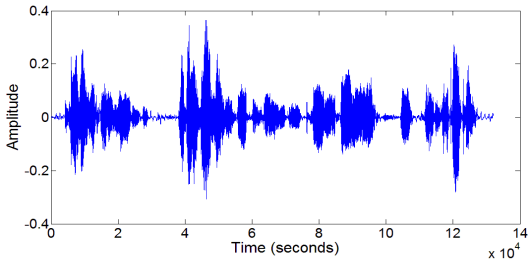
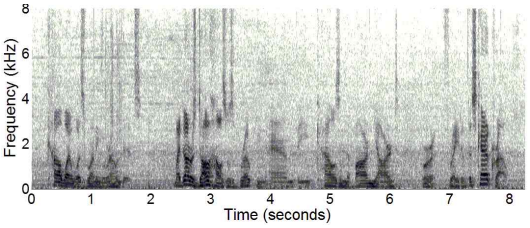
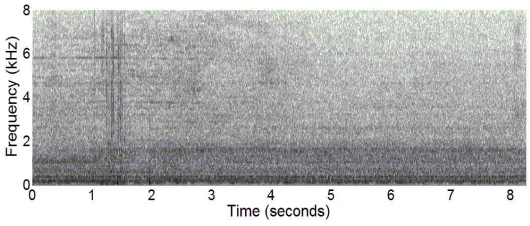


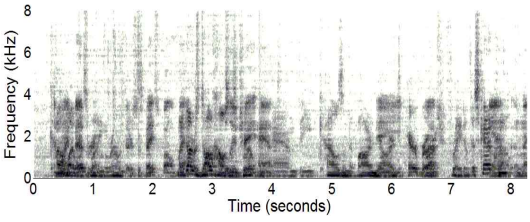
그림 4. 잡음이 중첩되지 않은 원래의 입력 음성
Fig. 4. Original input speech without degraded by noise



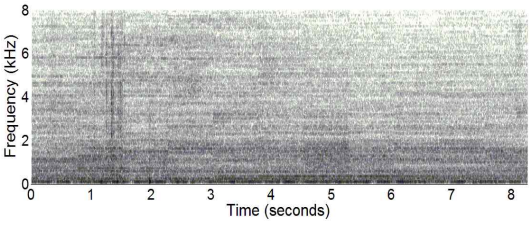
(a) 혼합 입력 음성신호
(a) Input mixed speech signal



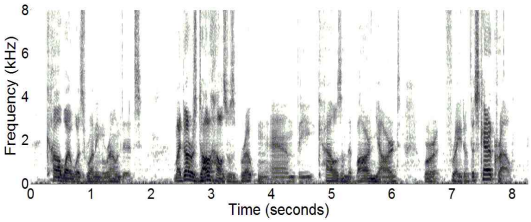
(b) 자동차 잡음
(b) Car noise



(c) 분리된 음성신호
(c) Separated speech signal



(d) 분리된 잡음신호
(d) Separated car noise



(e) 원래의 입력 음성신호
(e) Input clean speech
그림 5. 분리된 신호의 스펙트로그램
Fig. 5. Spectrogram for separated signal

그림 3의 IVA의 결과로부터 그림 4의 깨끗한 음성과 그림 3의 제안한 IVA에 의해 분리된 음성을 비교하였을 때 묵음구간의 잡음신호를 제외하고는 비교적 원 음원신호에 비하여 상당히 분리가 양호하게 된 것을 알 수 있다. 그리고 그림 3의 자동차 잡음에 대해서도 잡음성분을 거의 원 잡음신호에 가깝게 분리한 것을 실험을 통하여 확인할 수 있었다.

그림 5는 제안한 IVA 방법으로 처리한 분리된 신호의 스펙트로그램을 나타낸다. 그림 5(a)는 잡음이 중첩된 입력 혼합음성, 5(b)는 자동차 잡음, 5(c)는 IVA에 의해 분리된 음성, 5(d)는 IVA에 의해 분리된 잡음, 5(e)는 잡음이 없는 원래의 입력 음성신호를 각각 나타낸다. 그림의 결과로부터 제안한 IVA를 사용한 후의 스펙트럼이 원 음성 및 자동차 잡음과 비슷함을 알 수 있었다. 따라서 그림의 결과로부터 본 논문에서 제안한 방법은 충분히 음성과 잡음을 깨끗하게 분리할 수 있음을 확인하였다.

Ⅳ. 결 론

본 논문에서는 주파수영역의 실제 환경을 고려하여 음성 및 잡음의 두 음원에 대하여 주파수영역의 IVA를 실행하여 음원분리를 실시하였으며, 실험을 통해서 두 채널의 마이크로폰으로 입력된 혼합음성신호와 자동차 잡음을 분리하였다. 실험의 분리 결과로부터 분리된 음성신호는 묵음구간의 잡음을 제외하고는 원래의 음성신호와 비교하여 상당히 깨끗하게 분리가 된 것을 확인할 수 있었다. 그리고 잡음에 대해서도 잡음성분을 거의 원래의 잡음신호에

가깝게 분리한 것을 실험을 통하여 알 수 있었다. 또한 스펙트럼의 결과로부터 본 논문에서 제안한 IVA 방법을 사용한 후의 스펙트럼이 원 음성신호와 비슷함을 알 수 있었다. 따라서 본 논문에서 제안한 방법은 두 채널의 마이크로폰을 통해 음성신호에 잡음이 중첩되어 들어와도 음성과 잡음신호를 깨끗하게 분리할 수 있었다.

향후에는 실제 환경에 적용 가능하도록 자동차 잡음뿐만 아니라 다양한 잡음 및 다양한 음성데이터를 사용하여, 세 개 이상의 마이크로폰 신호를 분리하는 문제에 대하여 연구를 할 계획이다.

References

[1] S. Mirsamadi and J. H. L. Hansen, "Multichannel feature enhancement in distributed microphone arrays for robust distant speech recognition in smart rooms", IEEE Spoken Language Technology Workshop, pp. 507-512, Dec. 2014.

[2] J. H. Cho, "Efficient Compensation of Spectral Tilt for Speech Recognition in Noisy Environment", The Journal of IIBC, Vol. 17, No. 1, pp. 199-206, Feb. 2017.

[3] J. T. Oh, "A Study on the Design of Inaudible Acoustic Signal in Acoustic Communications and Positioning System", The Journal of IIBC, Vol. 17, No. 2, pp. 191-197, Apr. 2017.

[4] K. Nakadai, H. Nakajima, G. Ince, and Y. Hasegawa, "Sound source separation and automatic speech recognition for moving sources", IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 976-981, Oct. 2010.

[5] S. I. Kim, "Classification of Signals Segregated using ICA", Journal of the Institute of Electronics Engineers of Korea, Vol. 47, No. 4, pp. 10-17, Dec. 2010.

[6] K. S. Park, J. S. Park, K. S. Son, and H. T. Kim, "Postprocessing With Wiener Filtering Technique for Reducing Residual Crosstalk in Blind Source Separation", IEEE Signal Processing Letters, Vol. 13, No. 12, pp. 749-751, Dec. 2006.

[7] Z. Chu and K. S. Bae, "Post-processing of IVA-based 2-channel blind source separation for solving frequency bin permutation problem", Phonetics and Speech Sciences, Vol. 5, No. 4, pp. 211-216, Dec. 2013.

[8] X. Wang, X. Quan, and K. S. Bae, "Microphone Array Based Speech Enhancement Using Independent Vector Analysis", Phonetics and Speech Sciences, Vol. 4, No. 4, pp. 87-92, Dec. 2012.

[9] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model", Journal of the Acoustical Society of America, Vol. 124, No. 1, pp. 269-277, Jul. 2008.

[10] K. D. Donohue, "Systems Array Processing Toolbox [Online]", Available: http://www.engr.uky.edu/~donohue/, [accessed: Aug. 09. 2018].

[9] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", in Proc. ISCA ITRW ASR2000 on Automatic Speech Recognition: Challenges for the Next Millennium, Paris, France, Oct. 2000.

저자소개

최 재 승 (Jae-Seung Choi)

1989년 : 조선대학교 전자공학과 공학사

1995년 : 일본 오사카시립대학 전자정보공학부 공학석사

1999년 : 일본 오사카시립대학 전자정보공학부 공학박사

2000년 ~ 2001년 : 일본 마쯔시타 전기산업주식회사 (현, 파나소닉) AVC사 연구원

2002년 ~ 2007년 : 경북대학교 디지털기술연구소 책임연구원

2007년 ~ 현재 : 신라대학교 스마트전기전자공학부 교수

관심분야 : 음성 신호처리, 신경회로망, 필터링 및 잡음제거 등