



IVA에 의한 2채널 음원신호의 분리 및 반복스펙트럼차감에 의한 음성신호의 품질개선

최재승*

Separation of 2 Channel Source Signal by IVA and Improvement of Speech Signal Quality by Iterative Spectral Subtraction

Jae-Seung Choi*

요약

다채널의 혼합음성에 대하여 각 음원에 관한 정보를 추출하는 것은 인간의 음성신호처리 분야에 중요한 기술이다. 특히 혼합된 잡음음성 레벨이 회화음성 레벨보다 높게 되면 인간의 청각기강은 잡음음성을 청취하는데 곤란하게 되므로 잡음차감기술은 상당히 중요한 기술이다. 본 논문에서는 독립벡터분석 알고리즘에 의하여 잡음혼입 음성신호와 잡음신호를 분리한 후에 반복스펙트럼차감 알고리즘을 사용하여 잡음신호의 스펙트럼을 차감하는 알고리즘을 제안한다. 실험에서는 파형에 의한 실험결과와 SNR의 실험결과를 통하여 제안한 알고리즘을 검증하였다. 독립벡터분석에 의한 분리 실험결과 잡음혼입 음성신호와 잡음신호를 거의 원 신호에 가깝게 분리할 수 있었으며, 또한 반복스펙트럼차감 알고리즘을 사용함으로써 잡음혼입 음성신호는 깨끗한 음성신호의 입력 SNR과 비교하여 최대 10.41dB까지 개선되었다.

Abstract

In the field of human speech signal processing, it is an important technology to extract information about each speech source from multi-channel mixed speech signals. Especially, when the level of a mixed noisy speech signal is higher than the level of a conversational speech signal, the human auditory system becomes difficult to listen to the noisy speech signal, therefore the noise subtraction technique becomes a very important factor. This paper proposes an independent vector analysis (IVA) algorithm for separating the mixed noisy speech signal and the noise signal and an improved iteration spectral subtraction (IISS) algorithm for subtracting the spectrum of noise signals. In this experiment, the proposed algorithms were verified through experimental results of waveforms and SNRs. As a result of the separation experiment by IVA, it was possible to separate the mixed noisy signals and the noise signals close to original speech signals considerably. Moreover, by using IISS, the mixed noisy speech signals were improved up to 10.41dB compared with input SNRs of clean speech signals.

Keywords

independent vector analysis, separation of source, noise subtraction, mixed noisy speech

* 신라대학교 스마트전자공학부 교수
- ORCID: <http://orcid.org/0000-0002-5699-9701>

· Received: Jan. 26, 2019, Revised: Feb. 28, 2019, Accepted: Mar. 03, 2019

· 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

1. 서 론

최근 잡음환경에서의 음성인식장치는 마이크로폰에 근접된 위치의 음성에 대해서는 높은 음성인식 성능을 가지고 있는 반면에, 마이크로폰에서 멀리 떨어진 위치의 음성에 대해서는 주위 잡음 및 녹음실의 잔향의 영향을 받아서 음성인식 정도가 상당히 저하된다. 따라서 실제 환경에서 음성인식을 실시하기 위해서는 음성인식장치에 입력되는 음성에 대해서 특정한 처리를 실시 할 필요가 있다. 특히 컴퓨터에 의한 환경 중에서 여러 음성으로부터 관측된 혼합음성에 대하여 각각의 음원에 관한 정보를 추출하는 것은 로봇과 인간의 음성통신 분야에 있어서 중요한 기술 중의 하나이다[1]-[2].

잡음이 중첩된 실제 환경에 있어서 목적의 음원 신호만을 분리하기 위해서 마이크로폰을 사용한 Blind 신호분리(BSS, Blind Source Separation) 기술의 연구가 진행되고 있다[3]-[7]. BSS에서는 독립구성요소분석(ICA, Independent Component Analysis)[4]과 독립 Vector분석(IVA, Independent Vector Analysis)[5]-[7]에 기초한 수법이 현재 주목을 받고 있다. ICA는 주 구성요소 분석(PCA, Principal Component Analysis)과 동일하게 다차원신호의 해석에 이용하는 수법이다. PCA에서는 신호의 2차 상관관계만 주목하여 이 신호들을 무상관으로 변환하여 구하지만, ICA에서는 고차의 통계량 혹은 시간 상관계에 기초하여 독립성에 의하여 신호를 분리한다. ICA를 사용한 BSS는 각 음원의 통계적인 독립성을 가정함으로써 분리행렬을 추정하는 방법으로 선형혼합에 대해서는 유효한 수법이지만, 실제 환경에서의 잡음음성에 대해서는 그 유효성을 검증할 필요가 있다. 그러므로 본 논문에서는 먼저 잡음이 섞인 음성신호 및 잡음신호의 두 음원에 대한 주파수영역의 BSS의 실제 환경을 고려하여 IVA에 기초한 음원신호 분리를 목적으로 하며, 그 향상효과를 검증한다.

음성신호처리의 목적은 인간에게 정보전달을 하기 위한 수단으로 주로 사용된다. 그러나 잡음신호가 혼합된 음성이라도 인간의 음성처리기강은 충분히 정보신호를 전달하는데 문제는 없지만, 잡음혼입 음성신호 레벨이 회화음성신호 레벨보다 높게 되면 청취하는데 곤란하게 된다. 그러므로 잡음차감 기술

은 상당히 중요한 요소가 된다[8]-[10]. 따라서 본 논문에서는 먼저 제안한 IVA 알고리즘을 사용하여 음원신호를 분리한 후에 잡음차감 기술 중의 하나인 개량된 반복 스펙트럼차감(IISS, Improved Iteration Spectral Subtraction) 알고리즘을 제안한다. 즉, IVA에 의해 분리된 잡음혼입 음성신호에 대해서 IISS 알고리즘을 사용하여 잡음 스펙트럼을 차감하는 알고리즘을 제안한다.

제안한 IVA에 의한 알고리즘의 실험은 과형에 의해서 실험결과를 관찰하며, IISS에 의한 실험은 시간영역의 측정방법인 신호에 대한 잡음비율(SNR, Signal-to-Noise Ratio) 및 과형에 의해서 실험결과를 측정하여 제안한 알고리즘을 검증한다.

II. BSS에 기초한 IVA 알고리즘

BSS는 관측된 혼합 음성신호만을 사용하여 음원 신호를 추정하는 수법이며 몇 개의 마이크로폰에서 녹음된 혼합음성 등을 분리하는데 응용되고 있다[3]-[7]. 음성신호를 분리하는 경우에 시간차 및 레벨차를 다르게 하여 몇 개의 마이크로폰을 공간이 다른 위치에 배치하여 혼합한 음원신호를 만든다[11]-[12].

IVA[5]-[7]에 의하여 혼합신호를 분리할 경우에 두 개의 마이크로폰으로 입력된 서로 다른 음원을 분리한다면 음원신호 $s(k)$ 는 식 (1)과 같이 나타낼 수 있다.

$$s(k) = [s_1(k) \ s_2(k)]^T \quad (1)$$

여기에서 T 는 전치를 나타내고 k 는 이산시간을 나타내며 $s(k)$ 는 상호간에 독립되었다고 가정한다.

본 논문에서는 마이크로폰에서 수신된 부가잡음 $d(k)$ 로 손상된 음성신호 $s(k)$ 를 고려하여 혼합신호 $x(k)$ 를 $x(k) = s(k) + d(k)$ 로 나타내며, 두 개의 음원신호로부터 관측된 혼합신호 $x(k)$ 를 식 (2)와 같이 표현한다.

$$x(k) = As(k) + d(k) \quad (2)$$

여기에서 A 는 음원을 혼합신호로 변환하는 혼합행

렬, $x(k)$ 는 혼합신호 1과 2로 구성된다. 또한 A 는 식 (3)과 같이 나타낸다. IVA에서는 $s(k)$ 의 독립성의 가정으로부터 A 에 관한 지식을 사전에 가지지 않고 $x(k)$ 를 n 개의 독립적인 성분으로 분리하는 것이 목적이다.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (3)$$

식 (3)의 중첩된 혼합 신호는 식 (4)와 같이 표현한다.

$$x(k) = [x_1(k) \ x_2(k)]^T \quad (4)$$

IVA에 의하여 최종 분리된 신호는 음원신호 $y(k)$ 를 추정하기 위하여 혼합신호 $x(k)$ 와 분리행렬 W 를 사용하여 식 (5)의 $y(k)$ 를 생성한다.

$$y(k) = Wx(k) \quad (5)$$

여기서 $y(k)$ 는 $x(k)$ 의 추정치이며 $W=A^{-1}$ 이다. 식 (5)의 $y(k)$ 는 식 (6)의 분리된 음원 $y_1(k)$ 와 $y_2(k)$ 로 표현한다.

$$y(k) = [y_1(k) \ y_2(k)]^T \quad (6)$$

그러므로 W 를 식 (7)과 같이 나타낼 수 있다.

$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \quad (7)$$

III. 제안한 IISS 알고리즘

스펙트럼차감(Spectral Subtraction) 법은 Boll에 의해서 처음으로 제안되었으며, 스펙트럼영역에서 비음성 구간의 잡음을 추정하여 잡음이 섞인 음성신호의 스펙트럼에서 추정된 잡음 스펙트럼을 차감하는 방법이다[8].

본 논문에서는 개량된 반복 스펙트럼차감 알고리즘을 제안한다. 각 반복마다 형성된 비정상적인 잡음은 성질이 서로 다르기 때문에 IISS 알고리즘을 여러 차례 반복함으로써 잡음의 저감이 가능하다 [9]-[10]. IISS 알고리즘을 실행할 때에는 스펙트럼 차감 횟수를 10회 반복하였으며, 처음의 무음구간 9 프레임으로부터 각각의 잡음의 스펙트럼을 추정하여, 이 평균 스펙트럼을 잡음 스펙트럼으로 하였다. 그리고 분석 프레임마다 Fast Fourier 변환(FFT, Fast Fourier Transform)을 처리하여 잡음혼입 음성스펙트럼을 산출하며 주파수영역에서 잡음 차감 처리를 실행하였다. 또한 프레임의 추출은 해밍창으로 하였으며 한 프레임의 표본수는 512로 하였다.

그림 1은 본 논문에서 제안한 IVA 알고리즘 및 IISS 알고리즘의 개요를 나타낸다. 먼저 두 개의 마이크로폰으로 수신되는 잡음혼입 음성과 잡음의 두 개의 음원을 사용하며 각 음원은 서로 결합되어 각 마이크로폰에서 혼합 신호를 생성한다. 그리고 제안한 IVA에 의하여 음원을 추정하여 혼합신호를 분리한다.

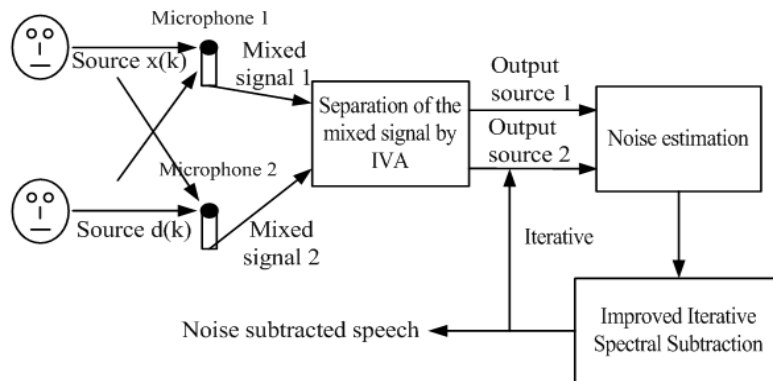


그림 1. 제안한 IVA 및 IISS 알고리즘
Fig. 1. Proposed IVA and IISS algorithms

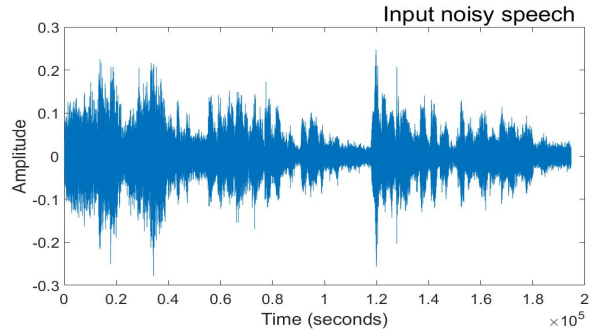
IVA에 의한 기본적인 분리방법은 FDICA (Frequency Domain Independent Component Analysis)를 이용하여 독립성을 기준으로 원신호를 분리한다. 다음으로 IVA에 의해서 분리된 잡음음성신호에 대하여 잡음추정을 한 다음에 제안한 IISS 알고리즘에 의하여 원래의 깨끗한 음성신호를 복원한다.

IV. 실험 결과

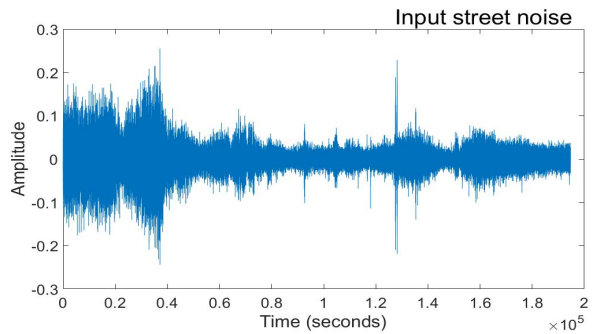
본 논문에서는 Image Source Method에 의해서 2채널의 혼합신호를 마이크로폰을 사용하여 합성하였다[11]. 실험에서 채택하는 혼합음원신호는 16kHz의 표본주파수로 수집한 남성화자 M1, M2, M3 및 여성화자 F1, F2, F3의 총 6개 음원신호를 사용하였으며, 음성신호의 길이는 평균 10초 정도이다 [12]. 음성에 중첩시키는 잡음신호는 Aurora2 database에 포함된 subway noise, airport noise, street noise, train noise의 4 종류를 사용하였다[13].

IVA의 분리 실험조건으로 분석 프레임 길이는 512 표본이며 75%의 중첩으로 각 프레임을 이동시킨다. 또한 분리행렬 W 를 추정하기 위하여 Step size $\eta = 0.1$, 최대 반복차수(Maximum Iterative Time)는 1,000회로 설정하였으며, 실험은 Windows 7 운영체제에서 Matlab(R2017a version) 프로그램을 사용하여 실험을 하였다.

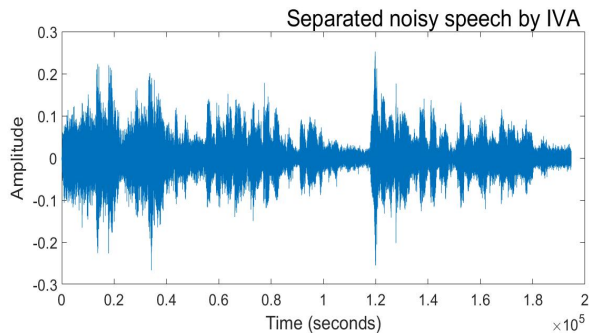
그림 2는 음성신호 F1에 street noise가 혼입 (SNR=3.45)된 입력 음성신호와 street noise에 대하여 제안한 IVA를 사용한 혼합신호의 파형분리 결과를 보여주고 있다. 그림 2(a)와 (b)는 각각 입력 잡음혼입신호 및 입력 잡음신호이며, 그림 2(c)와 (d)는 각각 IVA에 의해서 분리된 잡음음성신호 및 잡음신호이다. 그림의 결과로부터 제안한 IVA 알고리즘에 의하여 잡음이 중첩된 음성신호와 잡음신호를 각각 분리할 수 있었다. 특히 음성신호의 스펙트럼과 비슷한 street noise에 대해서 충분히 분리가 가능하다는 것을 파형의 결과로부터 확인된다. 실험결과에는 표시하지 않았지만 다른 3종류의 잡음에 대해서도 street noise와 비슷한 경향을 보여서 본 논문에서 제안한 IVA 알고리즘에 의해서 충분히 분리가 됨을 알 수 있었다.



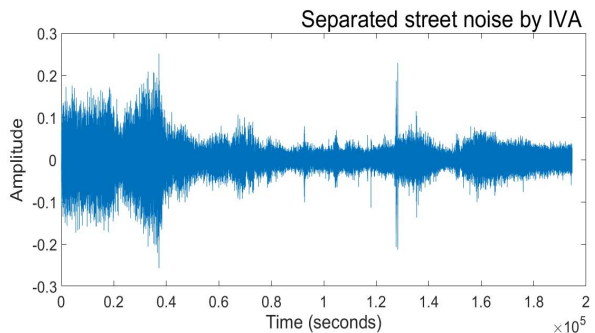
(a) 입력 잡음음성신호
(a) Input noisy speech signal



(b) 입력 street noise
(b) Input street noise



(c) 분리된 잡음음성신호
(c) Separated noisy speech signal



(d) 분리된 street noise
(d) Separated street noise

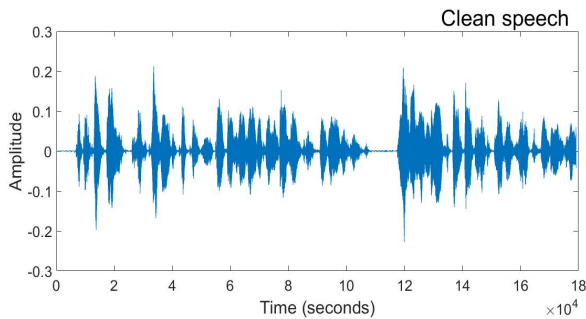
그림 2. 제안한 IVA에 의한 분리 실험결과
Fig. 2. Experiment results separated by proposed IVA

실험에서 사용한 음성품질의 평가법으로는 SNR을 사용하였다. SNR은 음성신호에 잡음이 어느 정도 추가되어 있는가를 나타내며 식 (8)과 같이 잡음 부가음성신호 $x(k)$ 로부터 음성신호 $s(k)$ 을 차감함으로써 SNR을 구한다.

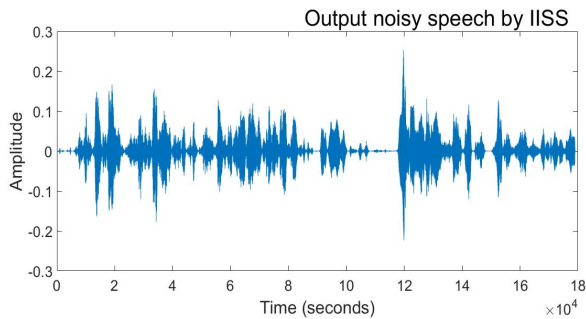
$$SNR = 10 \log_{10} \frac{\sum_{k=0}^{N-1} s(k)^2}{\sum_{k=0}^{N-1} [x(k) - s(k)]^2} \quad (8)$$

여기에서 N 은 데이터수를 나타내며, 산출된 SNR은 데시벨(dB) 단위로 구해진다.

그림 3은 음성신호 F1에 street noise가 혼입 (SNR=-3.45)된 음성신호와 street noise에 대하여 제안한 IISS를 사용한 잡음차감 알고리즘의 실험결과를 나타낸다. 그림 3(a)는 원래의 깨끗한 음성신호, 3(b)는 IISS에 의해 잡음이 차감된 출력신호를 각각 나타낸다.



(a) 깨끗한 음성신호
(a) Clean speech signal



(b) IISS에 의한 출력 잡음음성 신호
(b) Output noisy speech signal subtracted by IISS

그림 3. 제안한 IISS 실험결과
Fig. 3. Experiment results by proposed IISS

그림 2(c)의 IVA에 의해서 분리된 잡음음성신호와 비교하였을 때 IISS에 의해 잡음이 차감된 출력신호(그림 3(b))는 깨끗한 음성신호(그림 3(a))에 상당히 가깝게 원 신호로 복구됨이 확인된다.

제안한 IISS 알고리즘의 성능을 테스트하기 위하여 음성신호 F1에 subway noise, airport noise, street noise, train noise를 중첩시켜, 입력 SNR이 각각 -0.80dB, -3.23dB, -3.45dB, -5.87dB이 되도록 혼합음성신호를 생성하였다.

표 1은 4종류의 잡음이 중첩된 음성신호 F1에 대하여 제안한 IVA와 IISS의 출력 SNR을 각각 나타내고 있다. 표에서 Impr. 1은 입력 SNR과 비교한 개선된 IVA의 SNR이며, Impr. 2는 입력 SNR과 비교한 개선된 IISS의 SNR을 각각 나타낸다. 따라서 제안한 IISS 알고리즘을 사용함으로써 각 입력 SNR에 대하여 최대 10.41dB까지 개선됨을 검증하였다.

표 1. 출력 SNR의 성능 비교

Table 1. Performance comparison of output SNR

Noise	Input SNR (dB)	Output SNR (dB)			
		IVA	IISS	Impr. 1	Impr. 2
Subway	-0.80	6.66	7.98	7.46	8.78
Airport	-3.23	0.71	6.19	3.94	9.42
Street	-3.45	1.03	3.81	4.48	7.26
Train	-5.87	-2.64	4.54	3.23	10.41

V. 결 론

본 논문에서는 IVA 알고리즘에 의하여 잡음혼입 음성신호와 4종류의 각 잡음신호를 분리한 후에 IISS 알고리즘을 사용하여 잡음신호의 스펙트럼을 차감하는 알고리즘을 제안하였다. 제안한 알고리즘을 검증하기 위하여 파형에 의한 실험결과와 SNR의 실험결과를 통하여 제안한 알고리즘의 타당성을 검증하였다. 제안한 IVA 알고리즘에 의한 분리결과는 거의 원 신호에 가깝게 분리가 가능했으며, 또한 제안한 IISS에 의한 스펙트럼 차감결과는 원 신호의 입력 SNR에 대하여 최대 10.41dB까지 개선된 것을 확인할 수 있었다.

References

- [1] S. Nordebo, S. Nordholm, B. Bengtsson, and I. Claesson, "Noise reduction using an adaptive microphone array in a car-a speech recognition evaluation", Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 16-18, Oct. 1993.
- [2] J. Hong, S. Han, S. Jeong, and M. Hahn, "Adaptive microphone array processing for high-performance speech recognition in car environment", IEEE Transactions on Consumer Electronics, Vol. 57, No. 1, pp. 260-266, Feb. 2011.
- [3] V. Eksler, "Evaluation of Blind Separated Signals Using Speech Recognition System", EUROCON 2005-The International Conference on "Computer as a Tool", Vol. 2, pp. 1650-1653, Nov. 2005.
- [4] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals", IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 3, pp. 204-215, May 2003.
- [5] 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.
- [6] 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.
- [7] J. S. Choi, "A Blind Source Separation Method Based on Independent Vector Analysis for Separation of Speech Signal and Noise Signal", The Journal of Korean Institute of Information Technology, Vol. 16, No.10, pp. 69-74, Oct. 2018.
- [8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoustic, Speech and Signal Processing, Vol. 27, No. 7, pp. 113-120, Apr. 1979.
- [9] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise", IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 208-211, Apr. 1979.
- [10] O. Shinya and S. Tetsuya, "Iterative Spectral Subtraction Method for Reduction of Musical Noise", [in Japanese], Reports of the Meeting. the Acoustical Society of Japan, pp. 387-388, Mar. 2001.
- [11] 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.
- [12] K. D. Donohue, "Systems Array Processing Toolbox [Online]", Available: <http://www.engr.uky.edu/~donohue/>, [accessed: Aug. 09, 2018].
- [13] 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

최재승 (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 : 음성신호처리, 디지털신호처리, 음성인식, 마이크로폰 배열 음성신호처리