

주파수가중 스펙트럼성형필터와 위너필터를 결합한 음성 스펙트럼 강조

최재승*

Speech Spectrum Enhancement Combined with Frequency-weighted Spectrum Shaping Filter and Wiener Filter

Jae-Seung Choi*

Department of Electronic Engineering, Silla University, Busan 46958, Korea

요 약

디지털신호처리 분야에서 다양한 환경에서 존재하는 배경잡음을 제거하여 음성신호의 품질을 개선시키는 것은 반드시 필요한 문제이다. 음향학적으로 배경잡음을 제거할 때 고려해야 할 중요한 점은 인간의 청각기전이 주로 음성의 진폭 스펙트럼의 정보에 의존하여 문제 해결을 하고 있다는 사실이다. 본 논문에서는 음성의 진폭 스펙트럼의 추출을 주요 목적으로 하는 주파수가중 스펙트럼성형필터의 특성을 도입한다. 따라서 본 논문에서는 배경잡음으로 중첩된 음성신호 성분 중에 이 진폭 스펙트럼 정보를 추출하여, 위너 필터법과 음향학적인 모델에 의한 주파수가중 스펙트럼성형필터를 사용한 알고리즘을 제안한다. 본 실험에서는 스펙트럼 왜곡률(SD)에 의하여 제안한 알고리즘의 출력 SD가 기존의 다른 방법과 비교하여 약 5.28 dB 이상 개선되었다.

ABSTRACT

In the area of digital signal processing, it is necessary to improve the quality of the speech signal after removing the background noise which exists in a various real environments. The important thing to consider when removing the background noise acoustically is that to solve the problem, depending on the information of the human auditory mechanism is mainly the amplitude spectrum of the speech signal. This paper introduces the characteristics of a frequency-weighted spectrum shaping filter for the extraction of the amplitude spectrum of the speech signal with the primary purpose. Therefore, this paper proposes an algorithm using the methods of a Wiener filter and the frequency-weighted spectrum shaping filter according to the acoustic model, after extracted the amplitude spectral information in the noisy speech signal. The spectral distortion (SD) output of the proposed algorithm is experimentally improved more than 5.28 dB compared to a conventional method.

키워드 : 신호처리, 잡음, 진폭스펙트럼, 주파수가중 스펙트럼성형필터, 위너필터

Key word : Signal processing, noise, amplitude spectrum, frequency-weighted spectrum shaping filter, Wiener filter

Received 29 June 2016, Revised 29 June 2016, Accepted 15 July 2016

* Corresponding Author Jae-Seung Choi (E-mail:jschoi@silla.ac.kr, Tel:+82-51-999-5608)

Department of Electronic Engineering, Silla University, Busan 46958, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2016.20.10.1867>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

근년, 디지털신호처리 분야에서 다양한 환경에서 존재하는 배경잡음에 의해 오염된 음성신호로부터 이러한 배경잡음을 제거하여 음성신호의 품질 및 음성의 명료도를 개선시키는 것은 실제 환경의 다양한 분야에서 해결해야 할 반드시 필요한 문제이다. 일반적으로 배경잡음에 중첩된 음성신호에 사용되는 필터로서는 위너 필터(Wiener filter), 정합필터(Matched filter), 칼만필터(Kalman filter), 청각필터 등이 알려져 있다[1-4]. 그러나 음성신호에 배경잡음이 부가된 조건과 그 목적 등의 이유로부터 이러한 필터를 그대로 잡음 중의 음성신호에서 음성신호를 추출하고 필터링 처리하는데 사용하는 것은 고려해볼 필요가 있다[2, 3].

잡음 중의 음성신호 중에서 배경잡음을 제거한다는 것은 먼저 배경잡음의 특징 및 종류 등을 파악하는 것이 필요하다고 본다[5]. 배경잡음에는 여러 가지가 존재하고 있지만, 일반적으로 전력 스펙트럼을 미리 알고 있다고 가정하는 정상 가우시안잡음 및 실제 환경에서 많이 존재하는 많은 비정상잡음 등이 존재한다. 음성신호 처리 분야에서 청각필터 모델의 연구가 여러 가지 방법을 사용하여 많은 연구가 진행되고 있다. 이러한 청각필터 모델의 연구는 인간의 음향학적인 음성신호 청취에 관련한 분야의 공학적 응용에 공헌할 거라고 생각된다. 특히 적절하게 설계된 청각필터 모델은 실제의 인간 청각계에서 비선형적인 신호처리를 모의하고 있으므로 이 청각필터 모델의 출력은 음성신호의 특징을 스펙트럼으로 명확하게 표현할 수 있다. 효과적으로 음성인식 시스템의 특징 파라미터를 출력하는 청각필터 모델은 청각말초계의 청각기능 모델을 음성인식 시스템의 전처리로서 사용한 결과가 다수 보고되고 있으며, 또한 잡음환경 하에서 청각필터 모델을 사용하여 배경잡음을 제거하는 연구결과도 도출되고 있다[5-7].

음향학적으로 배경잡음을 제거할 때 고려해야 할 중요한 관점은 인간의 청각기전이 위상 스펙트럼의 정보에 민감하지 못하고 주로 진폭 스펙트럼의 정보에 의존하여 문제해결의 단서로서 인간의 음성을 지각하고 있다는 사실이다[6]. 따라서 본 논문에서는 인간의 청각기전의 특징 및 성질을 고려하여, 음성의 진폭 스펙트럼의 추출을 주요 목적으로 하는 음성의 주파수가중 스펙트럼성형필터의 특성을 도입한다. 이러한 필터 특성을

도입할 때에 문턱치를 설정하고, 잡음 성분으로 추정되는 문턱치 이하의 성분을 모두 제거하는 비선형적인 잡음제거를 실시한다. 또한 실제의 필터링을 실현하는데 있어서, 고속푸리에변환(Fast Fourier Transform, FFT)를 사용하여 주파수 영역에서 필터를 구성하여, 데이터 처리를 실시하는 방법을 사용한다.

본 논문의 실험에서는 음절 명료도 테스트를 하기 위하여 주관적인 평가로 알려진 주파수 영역에서의 스펙트럼 왜곡률(Spectral Distortion, SD)을 사용하여 실험 결과를 비교 고찰한다. 특히 본 논문에서 제안한 알고리즘을 사용함으로써 백색잡음에 대해서 약 5.28 dB 정도의 스펙트럼 왜곡률의 개선이 달성된 것을 4장에서 나타낸다.

본 논문은 크게 다음과 같은 구성을 한다. 2장에서는 본 논문에서 제안하는 알고리즘을 기술하고, 3장에서는 실험을 하는데 필요한 음성 및 잡음 데이터의 조건, 실험 평가방법 등을 기술한다. 4장에서는 음성 및 잡음 데이터를 사용하여 본 논문에서 제안한 알고리즘에 기초한 실험 결과 및 고찰을 나타낸다. 5장에서는 결론을 기술한다.

II. 제안한 알고리즘

일반적으로 음성신호를 청취하는 경우, 실제 환경에 존재하는 배경잡음에 의하여 출력 음성신호의 품질이 떨어져서 음성의 명료도가 떨어지는 경우가 가끔 생긴다. 이러한 문제를 해결하기 위하여 지금까지 스펙트럼 차감법, 위너필터법 등과 같은 다수의 음성강조수법 혹은 잡음차감법이 연구되어 왔다[8-13]. 이러한 방법들은 입력신호로부터 구해진 정보가 상당히 한정된다는 제약이 있기 때문에 사용하려는 배경잡음은 정상 혹은 준정상 상태 등의 제약조건을 설정할 필요가 있다. 따라서 제한된 정보 및 잡음의 통계적 특징만을 가지고 배경잡음을 완전히 제거하는 것은 어렵다고 판단된다. 그렇기 때문에 배경잡음에 대하여 미리 알고 있는 정보를 단서로 하여, 음성신호 성분의 어떠한 정보를 먼저 우선적으로 추출하는 것이 문제가 된다. 이때에 중요한 것은 1장에서 기술한 것과 같이 주로 진폭 스펙트럼의 정보를 단서로 하여 음성을 지각하고 있는 인간의 주파수가중 스펙트럼성형필터의 특성을 이용할 필요가 있

다. 그러므로 배경잡음에 중첩된 음성신호 성분 중에 이 진폭 스펙트럼 정보를 먼저 추출할 필요가 있다. 이하 본장에서는 배경잡음을 부가한 음성신호에 대하여 음향학적인 모델인 주파수가중 스펙트럼성형필터와 위너 필터법에 의한 알고리즘의 고찰을 실시한다.

음성신호 $s(k)$ 는 잡음신호 $n(k)$ 에 의해서 오염된다고 가정하면, 이 결과에 의해서 잡음에 오염된 음성신호 $y(k)$ 가 구해진다. 이 관계를 다음식과 같이 정의한다. 여기에서 k 는 시간을 나타내며, $n(k)$ 는 음성신호 $s(k)$ 와 비상관 관계가 있는 것으로 가정한다.

$$y(k) = s(k) + n(k) \quad (1)$$

본 논문에서는 인간의 청각기전의 특징을 응용하여 음성의 진폭 스펙트럼의 추출을 주요 목적으로 하는 청각가중 필터의 특성을 도입한다. 그림 1은 본 논문에서 제안하는 음향학적인 모델에 의한 주파수가중 스펙트럼성형필터(Frequency-weighted spectrum shaping filter)와 위너 필터법에 의한 알고리즘의 구성을 나타낸다. 제안한 알고리즘은 먼저 백색잡음으로 오염된 음성신호를 각 프레임에 대해 해밍창을 통과시킨 후에 잡음신호의 진폭 스펙트럼을 추정하기 위하여 음성신호가 결여된 프레임에서 잡음추정 방법을 채택한다[8]. 각 프레임에 대해서 선형예측 부호화(Linear Predictive Coding, LPC) 켈스트럼 계수의 음성특징 파라미터를 추출한 후에 본 논문에서 제안하는 잡음의 스펙트럼을 차감하는 위너필터에 의하여 백색잡음을 제거한다(그림 1의 좌측 블록). 여기에서 위너 필터법은 잡음이 중첩된 원래의 음성신호와 위너 필터에 의해 필터링된 신호와의 차이를 최소화하도록 설계한다[1, 9, 10]. 그림 1의 우측 블록에서 음성 데이터를 복소 고속푸리에변환(Fast Fourier Transform, FFT)을 한다. 다음으로 식 (1)을 사용하여 식 (2)의 전력 스펙트럼 $P_y(f)$ 를 구한다. 식 (2)에서 RE(f)와 IM(f)는 각각의 음성 데이터를 복소 FFT한 결과로부터 구해진 단시간 스펙트럼의 실수부와 허수부를 각각 나타낸다. 여기에서 f 는 주파수를 나타내며, $P_y(f)$ 는 $y(k)$ 의 전력 스펙트럼 밀도를 나타낸다.

$$P_y(f) = (RE(f))^2 + IM(f))^2 \quad (2)$$

본 논문의 실험에서 잡음 제거에 유용한 이동 진폭 평균 필터에 의한 3점 평균화를 이용하여 잡음 스펙트럼을 제거한다. 마지막으로 합성된 음성신호는 청각가중 필터에 의하여 합성된다.

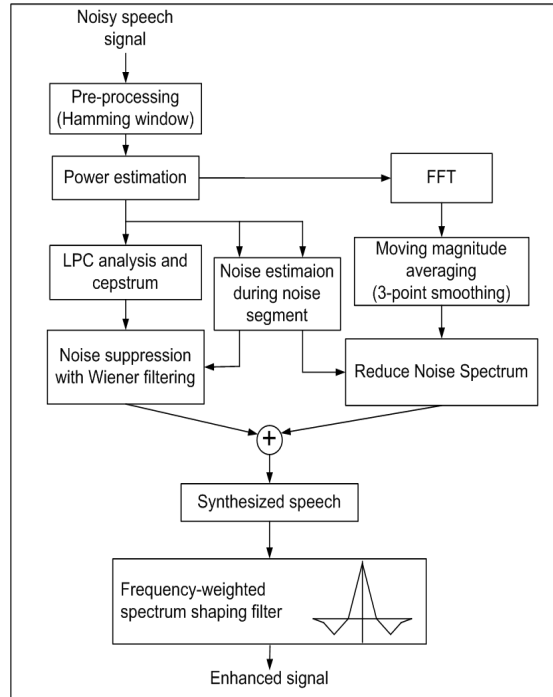


Fig. 1 Proposed algorithm using frequency weighted spectrum shaping filter and Wiener filter

신경생리학에서 측면 억제란 한 영역에 있는 신경 세포가 상호 간 연결되어 있을 때 어떤 신경 세포가 활성화되면서 주위의 신경세포는 억제되고, 가운데의 신경 세포는 활성화되는 형태로 나타난다. 이러한 측면억제에 의한 처리는 전처리로부터 출력되는 각 프레임의 스펙트럼에 대하여 주파수 축방향에 흥분성 및 억제성을 조합한 가중치 계수를 컨볼루션(Convolution)하여, 스펙트럼의 차이를 강조하는 처리를 한다. 이 측면억제 처리에 의해서 음성인식의 성능향상에 관한 연구 결과가 다수 보고되고 있다[4, 6, 7].

본 논문에서 사용한 측면억제 $Z(k)$ 의 처리는 식 (3)과 같이 나타낼 수 있다. 즉, $Z(k)$ 는 그림 2에 나타낸 음성가중 필터 $W(k)$ 와 합성된 음성신호 $Y(k)$ 의 컨볼루션 결과로 표현할 수 있으며, 최종적으로 입력된 진폭

스펙트럼 성분을 강조하여 배경잡음을 억제한다.

$$Z(k) = \sum_{i=-M}^{i=M} W(k-i) Y(k), k=0,1,\dots,N-1 \quad (3)$$

여기서 여기서 $Y(k)$ 는 잡음으로 오염된 음성신호의 FFT 출력의 진폭 성분이며, $W(k)$ 는 주파수 영역에서 흥분성 측면억제 및 억제성 측면억제의 성분을 조합한 가중치 계수를 나타낸다. 또한 N 은 한 프레임의 FFT 표본 수를 나타내며, 본 실험에서의 M 은 32로 한다.

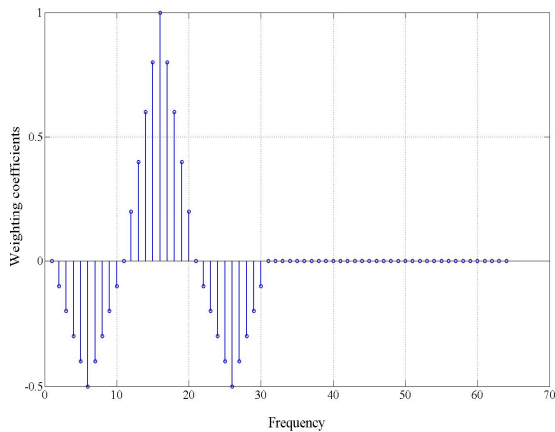


Fig. 2 Proposed speech weighting filter $W(k)$

III. 실험환경 및 실험 평가방법

본 논문에서는 배경잡음 억제 알고리즘 구현을 목표로 하여, 제안한 알고리즘을 이용하여 실험을 수행하였다. 실험환경은 Intel Core i7-2600K 3.80GHz CPU와 16GB RAM이 장착된 IBM 호환 컴퓨터에서 Windows 7 OS를 기반으로 하였으며, 기존의 상용 Microsoft VC 2008을 사용한 컴파일 환경을 이용하여 본 알고리즘을 테스트하였다.

본 실험에서 사용한 음성신호는 8 kHz의 샘플링 주파수의 음성 데이터를 사용하였으며, 일본인 단어 데이터베이스의 26 단어 중에서 임의적으로 선택하여 실험에 사용하였다. 본 실험에서는 임의적으로 선택한 남성 화자에 의한 단어("bizen(D1)", "shimonoseki(D2)")의 음성 데이터베이스를 사용하였다. 본 실험에서 사용한

잡음데이터는 컴퓨터의 시뮬레이션에 의해서 작성된 가우스 백색잡음(White noise, W)의 배경잡음을 사용하여 평가하였다.

본 논문에서 제안한 잡음제거의 평가척도로는 음성 명료도와 관계가 깊은 스펙트럼 왜곡률(Spectral distortion, SD)를 사용한다. 이 SD는 각 주파수 대역에서 입력신호의 대수 스펙트럼(dB)와 출력신호의 대수 스펙트럼(dB)의 차를 측정구간의 프레임 수에 대하여 구한다[7].

IV. 실험결과 및 고찰

본 논문에서 제안한 알고리즘은 3장에서 기술한 음성데이터와 백색잡음을 사용하여 잡음제거의 실험을 실시하였다.

먼저 제안한 알고리즘을 적용한 잡음제거 실험결과를 비교하기 위하여 백색잡음을 음성신호에 부가하였다. 그림 3은 실험에 사용된 원 음성신호 "D2"의 파형을 나타내고, 그림 4는 백색잡음이 음성신호에 부가된 잡음 음성신호의 파형을 나타낸다. 그림 5는 제안한 알고리즘을 사용할 때의 강조된 음성신호의 파형을 나타낸다. 따라서 이러한 그림들은 제안한 알고리즘을 사용할 때 배경잡음이 상당히 제거되는 것을 나타내고 있다.

표 1은 기존 방법과 차별성을 나타내기 위하여 본 논문에서 제안한 알고리즘과 기존의 청각필터[10]에 의한 방법을 비교한 실험결과이다. 표에서 백색잡음으로 오염된 음성신호의 입력 SD 값(Input SD values[dB])이 각각 20.77 dB, 18.76 dB일 경우에, 제안한 알고리즘(Prop. Alg.)의 출력 SD 값은 각각 10.79 dB, 10.56 dB 이 구해졌으며, 입력 SD 값과 비교하여 각각 9.98 dB, 8.20 dB(표 1의 Imp. Alg1.)이 개선된 것을 알 수 있다. 또한 제안한 알고리즘의 출력 SD 값은 기존의 청각필터 방법(표 1의 Aud. filter)보다 각각 5.28 dB, 4.41 dB(표 1의 Imp. Alg2.)이 개선되었다.

지금까지의 실험결과로부터, 본 논문에서 제안한 청각가중 필터에 의한 잡음제거를 각 프레임에서 실시함으로써 잡음이 많이 중첩된 음성에 대해서도 SD의 값을 경감하는 것을 알 수 있었다. 이상의 결과로부터 잡음이 중첩된 음성신호에 대해서도 본 방식에 의한 잡음제거의 효과가 있는 것을 알 수 있었다.

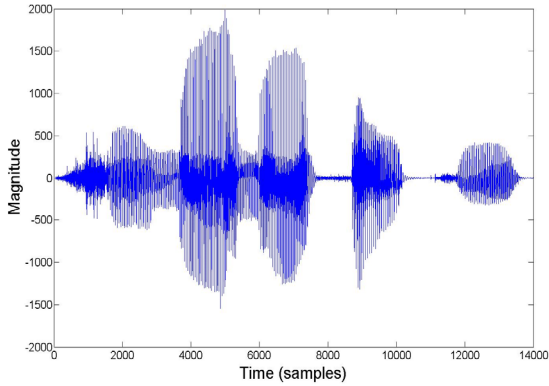


Fig. 3 The clean input speech signal

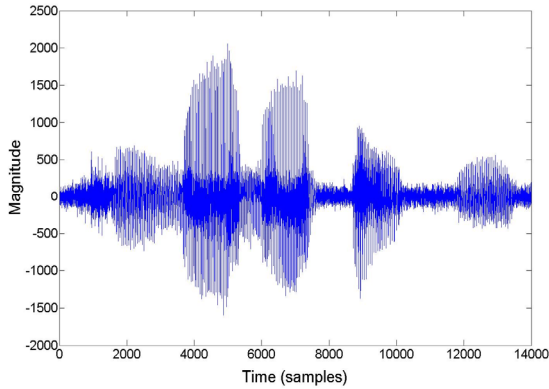


Fig. 4 The corrupted input speech signal with white noise

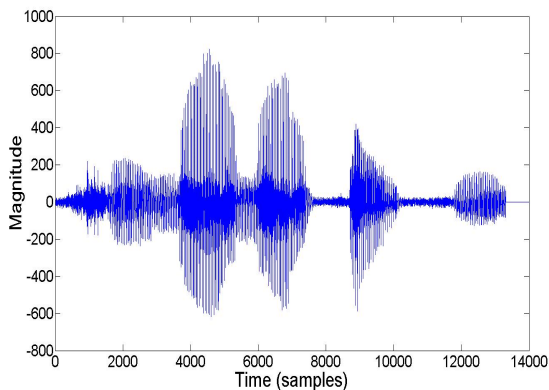


Fig. 5 The recovered output speech signal by the proposed algorithm

Table. 1 The comparison of SD values compared to a conventional method in the case of white noise

Noisy speech data	Input SD values [dB]	Output SD values [dB]			
		Aud. filter	Prop. Alg.	Imp. Alg1.	Imp Alg2.
D1+W	20.77	16.07	10.79	9.98	5.28
D2+W	18.76	14.97	10.56	8.20	4.41

V. 결 론

본 논문에서는 음성신호처리 시스템에 유용하게 사용되는 음성의 특징 파라미터를 출력하는 주파수가중 스펙트럼성형필터 모델과 위너필터를 사용하여, 잡음 환경 하에서 음성 중의 배경잡음을 제거하는 알고리즘을 제안하였다. 따라서 본 논문에서는 배경잡음을 제거할 때 중요하게 고려해야 할 인간의 청각기전이 음성의 진폭 스펙트럼의 추출을 주요 목적으로 하는 청각필터의 특성을 도입하였다.

본 논문에서 제안한 알고리즘은 기존의 청각필터와 비교하여 최대 5.28 dB 이상의 출력 SD 값이 개선된 것을 실험을 통하여 확인할 수 있었다. 또한 실험에 사용된 입력 음성신호와 제안된 알고리즘에 의한 출력 음성신호를 파형으로 비교한 결과 배경잡음이 명확하게 제거된 모양을 확인할 수 있었다. 이러한 결과로부터 본 논문에서 제안한 알고리즘이 잡음제거에 효과가 명확하다는 것을 알 수 있었다. 향후의 연구과제로는 여러 가지 비정상적인 잡음, 다양한 신호대잡음비를 사용하여 본 알고리즘의 성능향상 및 잡음제거를 검토할 필요가 있다고 본다. 본 연구를 더욱더 발전시킴으로써 청각실험에 도움이 된다고 기대한다.

REFERENCES

- [1] X. Dang and T. Nakai, "Noise reduction using modified phase spectra and Wiener Filter," *2011 IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1-5, September 18-21, 2011.
- [2] J. Freudenberger and S. Stenzel, "Blind Matched Filtering for Speech Recording in Uncorrelated Noise," *International Workshop on Acoustic Signal Enhancement*, pp. 1-4,

- September 4-6, 2012.
- [3] M. Mathe, S. P. Nandyala and T. K. Kumar, "Speech enhancement using Kalman Filter for white, random and color noise," *International Conference on Devices, Circuits and Systems (ICDCS)*, pp. 195-198, March 15-16, 2012.
- [4] J. H. L. Hansen and S. Nandkumar, "Robust Estimation of Speech in Noisy Backgrounds Based on Aspects of the Auditory Process," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3833-3849, June 1995.
- [5] M. F. R. Chowdhury, S. A. Selouani and D. O'Shaughnessy, "A highly non-stationary noise tracking and compensation algorithm, with applications to speech enhancement and on-line ASR," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4337-4340, March 25-30, 2012.
- [6] J. S. Choi, "Speech Processing System Using a Noise Reduction Neural Network Based on FFT Spectrums," *The Journal of Information and Communication Convergence Engineering*, vol. 10, no. 2, pp. 162-167, June 2012.
- [7] J. S. Choi, "Formant Enhancement Algorithm of Speech Using Auditory Filter," *The Journal of Korean Institute of Information Technology*, vol. 11, no. 7, pp. 173-178, July 2013.
- [8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.27, no.2, pp. 113-120, April 1979.
- [9] J. S. Choi, "A Wiener Filter Algorithm of Noise Subtraction Based on Threshold Detection," *The Journal of Korean Institute of Information Technology*, vol. 13, no. 7, pp. 51-56, July 2015.
- [10] J. S. Choi, "Noise Reduction Algorithm in Speech by Wiener Filter," *The Journal of Korea Institute of Electronic Communication Sciences*, vol. 8, no. 8, pp. 1293-1298, Sep. 2013.
- [11] N. Upadhyay, "An Improved Multi-band Speech Enhancement Utilizing Masking Properties of Human Hearing System," *Fifth International Symposium on Electronic System Design*, pp. 150-155, Dec. 15-17, 2014.
- [12] Md. T. Islam, C. Shahnaz and S. A. Fattah, "Speech enhancement based on a modified spectral subtraction method," *IEEE 57th International Midwest Symposium on Circuits and Systems*, pp. 1085-1088, Aug. 3-6, 2014.
- [13] S. Alaya, N. Zoghalmi and Z. Lachiri, "Adaptive filter for perceptual speech enhancement," *12th International Multi-Conference on Systems, Signals & Devices*, pp. 1-6, March 16-19, 2015.



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 주식회사) AVC사 연구원
2002년~2007년 경북대학교 디지털기술연구소 책임연구원
2007년~현재 신라대학교 전자공학과 교수
※ 주 관심분야 : 음성신호처리, 신경회로망, 적응필터와 잡음제거 등