Journal of KIIT. Vol. 16, No. 5, pp. 23-28, May 31, 2018. pISSN 1598-8619, eISSN 2093-7571 23
http://dx.doi.org/10.14801/jkiit.2018.16.5.23

음성강조를 위한 동시 마스킹 효과를 사용한 청각 필터링 알고리즘

최 재 승*

An Auditory Filtering Algorithm using Simultaneous Masking Effect for Speech Enhancement

Jae-Seung Choi*

요 약

음성신호처리 분야에서 인간의 청각 시스템은 잡음이 많은 환경에서도 우수한 성능을 발휘한다. 본 논문에서는 청각의 마스킹 효과를 주파수 영역에서 적용하여, 잡음을 억제하여 음성을 강조하는 동시 마스킹에 의한 청각 필터링 알고리즘을 제안한다. 제안하는 알고리즘은 상호억제 필터에 의한 청각 마스킹 모델, 스펙트럼 평활화 필터, MFCC 특징벡터를 적용함으로써 음성을 강조한다. 제안한 알고리즘은 주파수 영역의 스펙트럼을 강조함으로써 음성의 명료도를 향상시킨다. 실험의 성능 평가에는 Aurora2 음성 데이터베이스가 사용되며, 잡음환경으로는 백색 가우시안 잡음 및 자동차 잡음으로 오염된 음성신호와 5 종류의 SNR을 사용하여 실험을 수행한다. 실험 결과, 기존 방법과 비교하여 백색 가우시안 잡음에 대하여는 최대 5.3dB, 자동차 잡음에 대해서는 최대 4.9dB의 SNR이 향상되었다.

Abstract

In the field of speech signal processing, human auditory system excellently performs even in various noisy environments. This paper proposes an auditory filtering algorithm using simultaneous masking effect that enhances speech signal and reduces noise signal, by applying the masking effect of auditory system in frequency domain. The proposed algorithm enhances the speech signal by applying an auditory masking model using lateral inhibition filter, a spectrum smoothing filter, and MFCC feature vector. The proposed algorithm improves the intelligibility of the speech signal by enhancing the frequency domain spectrum. Aurora2 speech database is used for the performance evaluation of this experiment. Experiments are performed using five types of SNRs using noisy speech signals contaminated by white Gaussian noise and car noise. Experimental results show that segmental SNRs are improved by 5.3dB for white Gaussian noise and 4.9dB for car noise compared with conventional method.

Keywords
human auditory system, masking effect, lateral inhibition filter, speech enhancement, noise reduction

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

· Received: Jan. 18, 2018, Revised: Feb. 21, 2018, Accepted: Feb. 24, 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 휴대전화의 보급, 음성의 부호화, 디지털 음성신호처리 기술의 발전과 더불어 음성강조 기술들이 다양한 환경 및 기기에서 이용되고 있다. 이러한 음성강조 기술의 필요성에 따라서 휴대전화의 핸즈프리 기능, 인간과 컴퓨터와의 대화 방식 등과 같이 주위의 잡음에 의해서 음성품질의 열화가 문제가 되고 있다. 특히 음성통신을 수행하는 경우에 전달과정에서 잡음이 혼입되거나 왜곡이 발생함으로써 종종 신호대잡음비(SNR, Signal-to-Noise Ratio)가 저하하게 되어, 이러한 잡음이 음성통신에 있어서 큰 방해가 되고 있는 실정이다[1].

이러한 문제를 해결하기 위하여, 잡음에 의해서 품질이 저하된 음성을 개선하여 음성을 강조하는 여러 방식이 제안되어 있다. 예를 들면, 잡음에 강건한 특징량을 사용하는 수법, 음향 모델을 잡음환경에 적응시키는 수법, 위너필터, 적응필터 등의 여러 필터링에 의한 방법, 스펙트럼 상에서 잡음을 제거하는 스펙트럼 차감법(SS, Spectral Subtraction)은 닉마르코프모델(HMM, Hidden Markov Model)을 이용한 수법 들은 실제로 그 원리들이 다양하고 많은 개량 및 검토가 수행되어져 왔다[2]-[5]. SS 수법은 비음성 구간으로부터 잡음의 전력 스펙트럼을 추정한 후에 잡음이 부가된 음성의 전력 스펙트럼부터 추정된 잡음의 스펙트럼을 감산함으로써 잡음을 억압하는 수법이다[4]. 그러나 SS 수법은 정상적인 잡음을 가정하고 있기 때문에 비정상 잡음을 대상으로 하는 분야에서는 적용하기 어려운 문제점을 가지고 있다. 또한 잡음을 저감하기 위하여 음성에 전처리 및 후처리를 부가하는 방법도 제안되어 있으며, 여러 환경 및 상황에 따라서 유효한 방법을 찾고 있는 단계이다.

현재의 디지털 음성신호처리에 있어서 음성강조 및 잡음제거 방식은 중요한 기술이기 때문에 원래의 음성에 포함된 잡음을 제거하여 음성의 품질을 향상시킬 필요가 있다. 본 논문에서는 음성의 명료도를 개선하는 것을 목적으로 하여 잡음이 중첩된 음성에 대하여 음성강조처리를 수행하며, 이러한 목적을 위하여 백색 가우시안 잡음 (White Gaussian noise) 및 자동차잡음(Car noise) 환경 하에서 음성강조 알고리즘을 제안한다.

본 논문에서는 청각의 마스킹(Masking) 현상을 사용한 음성강조 알고리즘을 제안한다. 청각 마스킹이란 어떤 음에 대한 감지한도가 다른 음의 존재에 의해서 상승하는 현상이다. 이것은 청각의 주파수 선택기능에 의하여 신호와 마스커(Masker)를 분리 불가능할 때에도 발생하는 현상이다. 이 현상을 이용하여 음성의 마스킹 곡선에 의하여 낮은 범위의 잡음을 감쇄시킴으로써 스펙트럼을 차감하여 잡음을 억제하는 것이 가능하다. 따라서 본 논문에서는 청각의 주파수 선택기능에 의하여 신호와 잡음을 분리하는 청각의 마스킹현상을 사용한 음성강조의 수법을 제안한다.

본 논문에서는 잡음환경 하에 대하여 Aurora2 데이터베이스를 사용하여 실험을 수행하며, 종래 수법과의 성능비교를 통하여 제안한 알고리즘의 유효성을 확인한다. 본 실험에서는 객관적 성능 측정방법 중의 하나인 SNR을 사용하여 원음성과 처리 후의 강조음성에 대하여 성능평가를 실시한다.

Ⅱ. 청각의 마스킹 효과

마스킹은 청각이 가지는 기본적인 기능이며 다양한 청각의 현상에 관련되어 있으며, 일상생활 중에 주위로부터 발생하는 소음을 억제하는 대책의 하나이기도 하다. 이러한 마스킹의 특징은 음의 주파수 및 시간, 음의 크기 등의 요소에 의하여 마스킹되어 소음이 들리지 않게 되며, 두 개의 음을 동시에 청취할 때에 주파수가 분리되어 있다면 모든 음을 명료하게 청취할 수 있다[6][7].

마스킹 효과는 주파수 영역에서 발생하는 동시 마스킹(Simultaneous Masking)과 시간영역에서 발생하는 순시 마스킹(Temporal Masking)이 있다[7]. 본 논문에서는 청각의 동시적 마스킹 효과를 적용한 음성강조 수법을 사용한다. 이 마스킹 효과는 청각의 주파수 선택기능에 의하여 신호와 잡음을 분리 가능하도록 한다. 그림 1은 동시 마스킹 효과를 나타내며, 전방 마스크의 높은 음이 발생한 후에 일정 시간 동안의 후방 마스크의 낮은 음은 전방 마스크에 마스커 되어 들리지 않게 됨을 나타낸다.

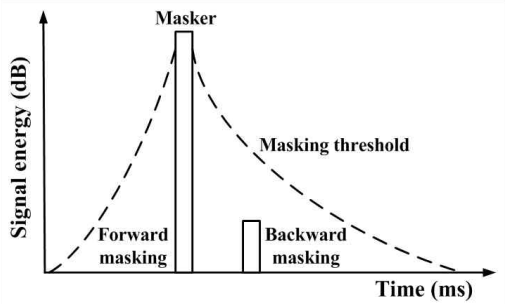

그림 1. 동시 마스킹 효과
Fig. 1. The effect of simultaneous masking

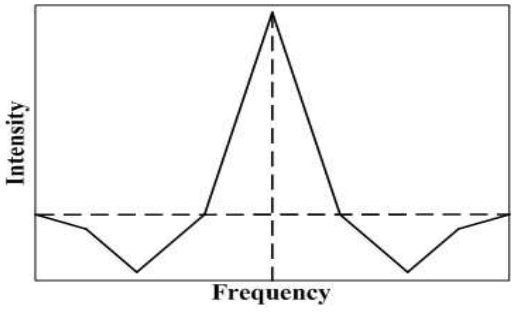

그림 2. 상호억제 필터의 특성곡선
Fig. 2. Characteristic curve of LIF

그림 2는 스펙트럼 영역에서의 잡음억제 효과를 나타내기 위하여, 음향 심리학에 관련된 상호억제 필터(Lateral Inhibition Filter, LIF)의 특성곡선을 나타낸다[8]-[10]. 이 곡선은 LIF 마스킹의 양이 주파수 영역에서 어떻게 변하는 지를 나타내며, 가로축은 주파수를 나타내고 세로축은 강도를 나타낸다. 이 곡선은 인간의 청각 시스템을 응용하여 주파수 영역에서 음성의 강도를 변화하게 한다.

III. 제안한 LIF 마스킹 알고리즘

동시 마스킹은 오디오 압축이나 음성강조 분야에서 보다 보편적이며, 음성강조를 위한 동시 마스킹의 여러 연구가 보고되었다. 동시 마스킹의 응용분야 중의 하나인 상호억제는 생리학, 신경생리학 실험에서 발견되었으며, 잡음 중의 음성에 대하여 잡음을 제거할 수 있다[7][8]. 상호억제는 주파수 영역에서의 각 프레임에 대하여 마스킹 효과로 작용하며, 이 기능은 입력의 스펙트럼을 날카롭게 하며 주변의 잡음을 감소시키는 기능을 가지고 있다[10].

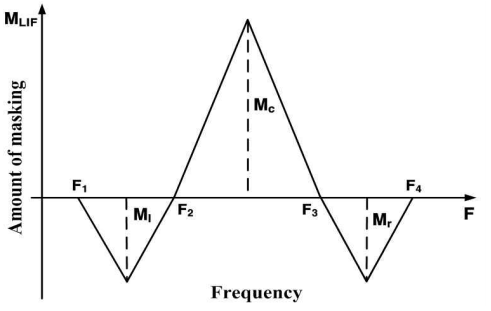

그림 3. 마스킹 처리를 위한 LIF의 청각필터 모델
Fig. 3. Acoustic filter model of LIF for masking

그림 3은 본 논문에서 제안하는 마스킹 효과가 적용된 청각필터의 모델을 나타낸다. LIF에 의한 파라미터는 주파수 영역에서의 음성의 스펙트럼을 날카롭게 하고 잡음을 경감하는 효과를 준다.

잡음이 중첩된 음성신호 $y(n)$은 식 (1)과 같이 음성 $s(n)$과 잡음 $d(n)$의 합으로 정의할 수 있다.

$$y(n) = s(n) + d(n) \tag{1}$$

식 (1)을 주파수 영역으로 변환하면 식 (2)와 같이 된다.

$$Y(\omega) = S(\omega) + D(\omega) \tag{2}$$

상호억제 마스커를 $M_{LIF}(\omega)$로 표현하였을 때, $M_{LIF}(\omega)$는 식 (3)과 같이 합의 평균값이 영이 되도록 제한 조건을 둔다. 이와 같은 제한은 잡음을 경감하는데 매우 유효하다.

$$\int_{-\infty}^{\infty} M_{LIF}(\omega)\, d\omega = 0 \tag{3}$$

따라서 본 실험에서는 주파수변화량을 나타내는 마스킹 필터 길이를 $\Delta F = F_2 - F_1 = F_3 - F_2 = F_4 - F_3 = 10$과 같이 실험에 의해서 최적의 파라미터를 설정한다. 식 (4)는 $M_{LIF}(\omega)$의 최종 필터 계수를 나타낸다.

$$M_l + M_c + M_r = 0 \tag{4}$$
$$M_c = 1,\ M_l = M_r = -0.5$$

상호억제에 의해서 마스킹 효과가 적용된 (i) 번째 프레임의 스펙트럼 $\hat{P}_y^{(i)}(\omega)$는 식 (5)와 같이 고속 푸리에 변환(FFT, Fast Fourier Transform)된 잡음으로 오염된 음성신호의 주파수 스펙트럼 $P_y(\omega)$와 상호억제 마스커 $M_{LIF}(\omega)$이 합성된 음성신호의 컨볼루션(Convolution) 처리에 의하여 구해진다. 따라서 마스킹 효과에 의하여 최종적으로 출력된 주파수 스펙트럼 성분을 강조하여 배경잡음을 억제한다.

$$\hat{P}_y^{(i)}(\omega) = \int_{-\infty}^{\infty} P_y^{(i)}(\omega)\, M_{LIF}(\omega)\, d\omega \tag{5}$$

단시간 전력 스펙트럼은 음성의 스펙트럼 피크 및 분석창의 부엽(Sidelobe)에 의한 피크를 포함하기 때문에 잡음 전력 스펙트럼은 프레임 사이에서 불규칙한 변동을 발생시킨다. 따라서 각 프레임에서 원하지 않은 불규칙적인 잡음을 제거하기 위하여 스펙트럼 평활화 필터(SSF, Spectrum Smoothing Filter) $\tilde{P}_{ssf}^{(i)}(\omega)$를 식 (6)과 같이 정의한다.

$$\tilde{P}_{ssf}^{(i)}(\omega) = \frac{1}{2N_{ssf}+1} \sum_{j=-N_{ssf}}^{N_{ssf}} \beta_m \hat{P}_y^{(i-j)}(\omega) \tag{6}$$

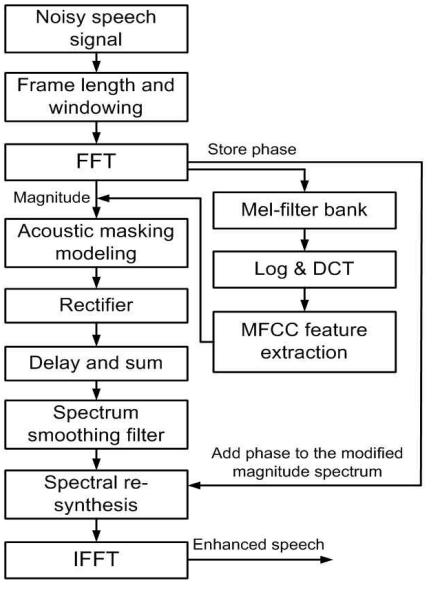

그림 4. 제안한 청각 마스킹 알고리즘
Fig. 4. Proposed acoustic masking algorithm

본 실험에서는 $N_{ssf}=2$, $[\beta_{-2}\beta_{-1}\beta_0\beta_1\beta_{-2}]$=[0.7, 1.1, 1.4, 1.1, 0.7]와 같이 설정한다.

그림 4는 본 논문에서 제안한 상호억제의 마스킹 효과가 적용된 전체 시스템 블록도를 나타낸다. 8 kHz로 표본화된 잡음으로 오염된 음성신호는 256 표본값(32ms)의 해밍창이 곱해진 후에 FFT에 의해서 스펙트럼 성분이 구해진다. 이 후에 14차의 MFCC 켑스트럼 계수가 계산된 다음에 상호억제에 의한 마스킹 효과가 적용된다. 상호억제 마스킹 이후의 출력에 포함된 음의 값은 모두 가청 임계 값 이하임을 의미하기 때문에 인간의 청각 시스템을 기반으로 하는 마스킹 효과에 있어서는 가청 임계 값 이하의 모든 신호는 청각처리에서 무의미하다. 따라서 모든 음의 출력을 0으로 설정하기 위해 상호억제 마스킹 후 정류기가 사용된다. 스펙트럼 평활화 필터를 적용하기 위하여 프레임 단위로 3 프레임 지연이 발생하여 가중치 합에 의한 스펙트럼 평활화 및 포르만트 강조가 된다. 마지막으로 위상 스펙트럼이 마스킹 처리된 음성신호와 합성된 후에 고속 역 푸리에 변환(IFFT, Inverse Fast Fourier Transform)됨으로써 강조된 음성신호를 구할 수 있다.

Ⅳ. 실험 조건 및 결과

본 논문에서는 8kHz로 샘플링된 Aurora2의 연속 영어숫자 음성데이터를 사용하여 잡음환경 하에서 제안한 알고리즘의 평가실험을 수행하였다[11]. Aurora2 데이터베이스의 평가 셋트에는 A, B, C 셋트와 같이 3 종류의 서로 다른 음성데이터로 분류되어 있다. 각 셋트에는 성인남성 52명, 성인여성 52명에 의한 합계 4004 발성문의 음성이 있으며, 이것을 4개로 분할한 1001 발성이 기본단위로 되어 있다. 여기에는 서로 다른 잡음을 중첩하여 7 종류의 SNR(-5, 0, 5, 10, 15, 20, ∞ [dB])로 분류한다.

본 논문에서는 제안한 알고리즘의 평가를 위하여 객관적 성능측정 방법인 입력 SNR(SNR_I)과 출력 SNR(SNR_O)을 사용하였다.

$$SNR_I = 10\log_{10}\frac{\sum_{i=0}^{M-1} s^2(n)}{\sum_{i=0}^{M-1} d^2(n)}\ [dB] \tag{7}$$

$$SNR_O = 10\log_{10}\frac{\sum_{i=0}^{M-1} s^2(n)}{\sum_{i=0}^{M-1}[s(n)-\hat{s}(n)]^2}\ [dB] \qquad (8)$$

여기에서 M은 프레임 길이이며 본 실험에서는 M=256이다. F는 음성신호의 전체 프레임 수이며, $\hat{s}(n)$은 강조된 시간영역의 출력 음성신호이다.

본 실험에서는 A 셋트에 포함된 30명의 음성데이터와 백색 가우시안 잡음 및 자동차잡음에 대하여, 다양한 SNR 잡음레벨(SNR_I=20, 15, 10, 5, 0 [dB])을 사용하여 본 논문에서 제안한 알고리즘의 성능 테스트를 하였다.

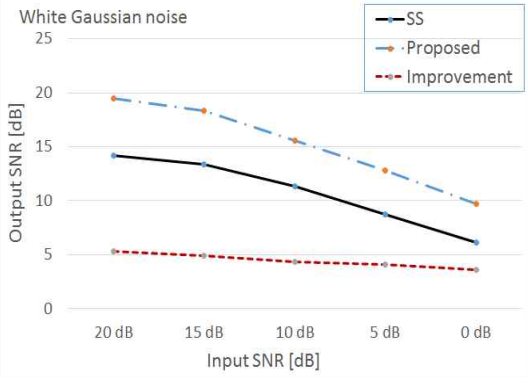

그림 5. 백색잡음에 대한 SNR 개선의 성능비교
Fig. 5. Performance comparison of SNR improvement in Gaussian white noise

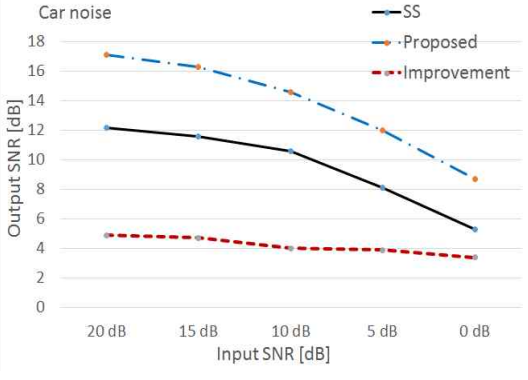

그림 6. 자동차잡음에 대한 SNR 개선의 성능비교
Fig. 6. Performance comparison of SNR improvement in car noise

그림 5와 그림 6은 본 논문에서 제안한 알고리즘과 기존의 SS 수법[4]의 SNR_O에 의한 성능개선의 실험결과를 비교하였다. 그림 5는 백색 가우시안 잡음에 대한 실험결과이며, 입력 SNR에 대하여 최대 5.3dB의 출력 SNR_O이 개선되었다. 그림 6은 자동차잡음에 대하여 최대 4.9dB의 출력 SNR_O이 개선된 것을 알 수 있다.

실험결과로부터 백색 가우시안 잡음은 물론 유색잡음인 자동차잡음으로 오염된 음성신호에 대해서도 본 논문에서 제안한 알고리즘이 유효한 것을 알 수 있었으며, 잡음이 높았을 때에도 SNR_O 결과가 양호한 것을 확인할 수 있었다. 특히 본 논문에서 제안한 알고리즘의 성능개선은 크다고 할 수 있으며, 미지의 음성 및 잡음 환경 하에서도 충분한 성능을 발휘하였다.

V. 결 론

본 논문에서는 청각 마스킹효과, 스펙트럼 평활화, MFCC 특징벡터를 사용한 음성강조 알고리즘을 제안하였다. 제안한 알고리즘은 청각의 주파수 선택기능에 의하여 신호와 잡음의 분리가 가능한 마스킹 효과에 의한 음성강조 수법이며, 2 종류의 잡음 및 5 종류의 SNR 환경 하에서 Aurora2 음성 데이터베이스를 사용하여 실험을 수행하였다. 기존의 수법과의 성능비교를 통하여 제안한 알고리즘이 유효하다는 것을 알 수 있었으며, 백색 가우시안 잡음에 대해서는 SNR_O이 최대 5.3dB 향상되었으며, 자동차잡음에 대해서는 최대 4.9dB의 SNR_O이 향상된 것을 알 수 있었다. 결론적으로 본 논문에서 제안한 알고리즘이 낮은 SNR 환경 하에서도 상당히 효과적이라는 것을 확인할 수 있었다.

References

[1] S. Bavkar and S. Sahare, "PCA based single channel speech enhancement method for highly noisy environment", International Conference on Advances in Computing, Communications and Informatics, pp. 1103-1107, Aug. 2013.

[2] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, pp. 1218-1234, Jul. 2006.

[3] J. Du, Y. Hu, L. R. Dai, and R. H. Wang, "HMM-based pseudo-clean speech synthesis for splice algorithm", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4570-4573, Mar. 2010.

[4] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 27, No. 2, pp. 113-120, Apr. 1979.

[5] J. S. Choi, "Speech-dependent Speaker Identification Using Mel Frequency Cepstrum Coefficients for Continuous Speech Recognition", Journal of KIIT, Vol. 14, No. 10, pp. 67-72, Oct. 2016.

[6] B. Li and K. C. Sim, "A Spectral Masking Approach to Noise-Robust Speech Recognition Using Deep Neural Networks", IEEE/ACM Trans. on Audio, Speech, and Language Processing, pp. 1296-1305, Vol. 22, No. 8, Jun. 2014.

[7] P. Dai and I. Y. Soon, "An improved model of masking effects for robust speech recognition system", Speech Communication, Vol. 55, No. 3, pp. 387-396, Mar. 2013.

[8] Y. M. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence", IEEE Transactions on Signal Processing, Vol. 39, No. 9, pp. 1943-1954, Sep. 1991.

[9] T. Houtgast, "Psychophysical evidence for lateral inhibition in hearing", The Journal of the Acoustical Society of America, Vol. 51, No. 6B, pp. 1885-1894, Jun. 1972.

[10] J. S. Choi, "An Adaptive Speech Enhancement System Based on Noise Level Estimation and Lateral Inhibition", Acta Acustica united with Acustica, Vol. 93, No. 4, pp. 507-673, July/Aug. 2007.

[11] 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 주식회사) AVC사 연구원

2002년 ~ 2007년 : 경북대학교 디지털기술연구소 책임연구원

2007년 ~ 현재 : 신라대학교 스마트전기전자공학부 교수

관심분야 : 음성신호처리, 신경회로망, 적응필터와 잡음제거, 디지털 TV 및 멀티미디어 등