진폭 및 위상스펙트럼이 도입된 신경회로망에 의한 잡음억제 알고리즘

최 재 승*

Noise Suppression Algorithm using Neural Network based Amplitude and Phase Spectrum

Jae Seung Choi*

요 약

본 논문에서는 다양한 배경잡음에 의해 열화된 음성을 강조하기 위하여 청각모델에 기초로 한 잡음환경에 적응된 잡음억제 시스템을 제안한다. 제안한 시스템은 먼저 유성음, 무성음 및 묵음의 구간을 검출한 후, 각 입력 프레임에서 적응적인 청각기강의 처리를 한다. 마지막으로 진폭성분과 위상성분이 포함된 신경회로망을 사용하여 잡음신호를 제거한 후에 음성을 강조하는 처리를 한다. 본 시스템은 신호대잡음비의 평가방법을 통하여 다양한 잡음에 의해서 열화된 음성신호에 대해서 유효하다는 것을 실험으로 확인한다.

ABSTRACT

This paper proposes an adaptive noise suppression system based on human auditory model to enhance speech signal that is degraded by various background noises. The proposed system detects voiced, unvoiced and silence sections for each frame and implements an adaptive auditory process, then reduces the noise speech signal using a neural network including amplitude component and phase component. Based on measuring signal-to-noise ratios, experiments confirm that the proposed system is effective for speech signal that is degraded by various noises.

키워드

Adaptive noise suppression, human auditory model, amplitude and phase spectrum, neural network

I. 서 론

근대과학의 진보에 따라서 정보처리에 관한 분야가 눈부시게 발전하고 있다. 이러한 분야 중에서 통신신호처리, 음성인식, 인공지능 등의 연구가 최근 활발히 진행되고 있으며 신세대 컴퓨터 등의 분야도 주목받고 있다. 이 중에서 중요한 문제 중의 하나로서 음성인식 등의 음성정보처리의 실용화를 위해서 실제 환경에 있어서 배경잡음에 대한 대응이 중요시되고 있다.

과거에 개발된 잡음억제시스템은 음성 혹은 잡음의 성질에 기초를 두었다. 적응적인 잡음 제거[1, 2]는 음성 혹은 잡음에 상관이 있는 적응적으로 가중치를 적용한 참조신호에 의해서 최소 2승평균을 사용하여 잡음을 감소시켜 음성신호를 강조한다. 적당한 데이터

* 신라대학교 전자공학과

접수일자 2008. 11. 07

베이스에 기초한 강조함수를 학습하는 데는 신경회로망(Neural Network: NN)을 사용한다[3]. Spectral subtraction[4, 5]에서는 음성신호가 잡음과 무상관이라고 가정하여, 강조된 음성의 진폭스펙트럼은 잡음을 포함한 음성의 비음성의 활동범위에서 추정된 잡음의 스펙트럼을 제거하여 구해진다. 강조된 음성은 이렇게 해서 구해진 진폭스펙트럼과 원래의 위상스펙트럼으로부터 역푸리에변환(Inverse fast Fourier transform : IFFT)에 의해서 재구성된다. 인간의 청각시스템은 배경잡음을 압축하는 것이 가능하고, 음성과 잡음에 대한 사전의 지식없이 희망하는 신호를 선택이 가능하다[6, 7]. Ghiza[7]는 완전한 와우각(달팽이관)모델을 사용하여, 음성신호는 청각의 처리를 통하여 강조되는 것을 나타내고 있다. 본 연구에서는 상호억제(lateral inhibition)라고 불리우는 하나의 청각모델을 사용하여 연구한다. 이것은 생리학 및 심리학을 통해서 발견되어 신경생리학은 이것을 음성의 스펙트럼을 날카롭게 하는 것으로부터 관계되었다.

일반적으로 입력음성신호를 고속푸리에변환(fast Fourier transform : FFT)하여 스펙트럼의 진폭성분과 위상성분으로 분리하여 진폭성분만을 조작하여 음성신호처리를 한 후에 위상성분은 그대로 사용하고 있다. 이것은 위상성분보다 진폭성분이 음성정보를 많이 포함하고 있기 때문이다. 그러나 이러한 처리에 의하여 진폭성분과 위상성분의 물리적인 부정합에 따라서 악음적 잡음의 원인을 일으키게 되며, 이것을 해결하지 않는 한 강조하였다고 생각한 목적 신호가 오히려 듣기 어려워지는 원인을 제공한다. 따라서 본 논문에서는 신경회로망을 이용하여 음성신호를 처리하는 경우에도 시간성분뿐만 아니라 위상성분도 신경회로망에 포함시켜 학습시킴으로써 1) 인간의 귀의 구조에 근접한 신호처리가 가능하게 하는 것, 2) 특정한 음성을 식별하여 인식 가능하게 하는 것, 3) 악음적 잡음의 제거를 통하여 잡음이 중첩된 음성을 명확하게 인식하여 음성처리를 한다.

본 논문에서는 인간의 귀에 대한 신호처리를 통하여 잡음이 존재하는 환경 하에서 먼저 상호억제라고 하는 청각기강을 공학적으로 응용하는 방법을 제안하며, 진폭성분 NN와 위상성분 NN로 구성된 NN 시스템을 적용하여, FFT한 진폭성분 및 위상성분을 복원하는 알고리즘을 제안한다. 그리고 유성음과 무성음의 구간을 검출한 후, 저역, 중역, 고역으로 분리된 NN을 사용하여 잡음신호를 제거한 후에 음성을 강조하는 처리를 한다.

II. 제안한 신경회로망 시스템

계산기에 의한 청각특성의 분석에 있어서, 음성 및 음악 등의 정상적인 배경잡음을 억제하는 것은 목적으로 하는 신호를 고정도로 추출하기 위해서도 중요하다. 이와 같은 목적으로 사용가능한 잡음억제처리로 가장 먼저 인용되어지는 방식은 Boll에 의한 스펙트럼 차감법[5]이다. 그러나 이 방법은 청취자를 상정한 분석 합성계를 사용하는 경우에는 그다지 사용하고 있지 않다. 이것은 합성음에 악음적 잡음이 발생하기 때문에 강조하였다고 생각한 목적 신호가 도리어 듣기 어려워지기 때문이다. 또한 다른 청각기능 분석시스템의 전처리로서 사용하는 경우에도 예측 및 제어를 할 수 없는 악음적 잡음을 발생시키지 않는 경우가 좋다. 이 악음적 잡음의 결점을 극복하기 위해서는 몇 가지 새로운 개량방법이 제안되어 있지만[8, 9], 신호대잡음비(Signal-to-Noise Ration; SNR)가 0 dB에 가까운 경우의 SNR 개선도의 평가에 대한 결과는 없으며, 항상 유효한 방법이 되는가에 대해서도 정확히 알 수 없다. 원래 스펙트럼 차감법에서는 스펙트럼의 진폭성분과 위상성분을 분리하여 진폭성분만을 조작하여 위상성분은 그대로 사용하고 있다. 이 처리에 의한 진폭 및 위상성분의 물리적인 부정합이 악음적 잡음의 원인이 되며, 이것을 해소하지 않는 한 본질적인 해결을 할 수 없다. 그림 1은 입력층, 중간층, 출력층으로 구성된 퍼셉트론형의 일반적인 신경회로망이다. 본 논문에서는 오차역전파방식에 의해 네트워크를 학습시키며, 이 방식은 입력된 학습패턴에 대하여 학습데이터와 출력과의 오차의 절대 값의 합이 구해져 이 오차가 일정한 값보다 적게 되도록 결합계수가 변경된다. 유닛의 입출력함수는 비선형 함수를 사용하고 결합계수의 수정에는 가속도계수를 사용하는 모멘트법을 채용하였다.

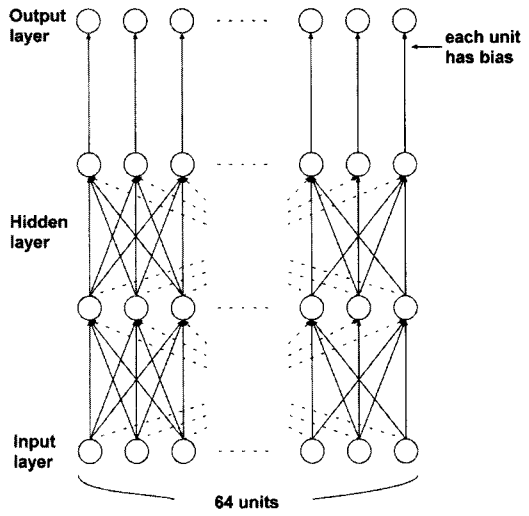

그림 1. 신경회로망의 구성
Fig. 1 Construction of neural network.

본 논문에서는 이러한 악음적잡음이 발생하는 문제점을 본질적으로 해결하기 위하여 잡음에 적응적인 잡음억제 알고리즘을 제안한다. 이 방법은 스펙트럼 차감법과 동일한 전제 조건으로 동등의 SNR 개선이 가능하며, 또한 악음적 잡음을 발생시키지 않는 방법이다. 따라서 본 논문에서는 잡음억제를 위한 시불변의 단시간 푸리에 변환뿐만 아니라 진폭 및 위상성분이 도입된 그림 2의 신경회로망 시스템을 제안한다.

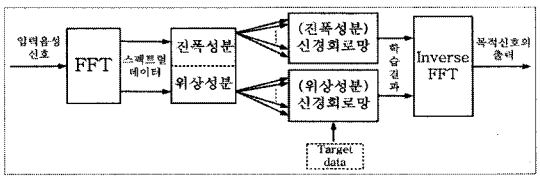

그림 2. 제안한 신경회로망 시스템
Fig. 2 Proposed neural network system.

제안한 신경회로망 시스템은 128샘플(16 ms)의 입력 데이터에 대해서 FFT를 실시하여 스펙트럼 데이터를 구한다. FFT에 의해 구해진 진폭 및 위상성분은 FFT의 63번째의 성분을 중심으로 대칭한 값을 가지므로, 용장부를 제외한 각 64샘플(0~63샘플)의 진폭 및 위상성분을 각각의 NN의 입력으로 한다. 따라서 스펙트럼 데이터를 각각 64개의 입력층 유닛, 64개의 중간층 유닛, 64개의 출력층 유닛으로 구성된 3층의 NN에 입력함으로써, 각 출력신호는 학습신호와 일치한 정확한 값을 취하도록 학습한다. 본 실험에서는 초기 가중치를 -0.05~+0.05의 난수를 사용하였으며, 학습계수 $\alpha = 0.1$, 가속도 계수 $\beta = 0.6$로 하여 최대 학습횟수를 평균2승 오차의 변화가 거의 없어지는(0.0001 이하) 15,000회로 하였다.

Ⅲ. 인간의 청각 모델

생물계의 감각 수용기에서 발견된 상호억제는 인간의 청각기능을 모델로 한 청각시스템이며, 배경잡음을 압축하는 것이 가능하고 음성과 잡음에 대한 사전의 지식없이 희망하는 신호의 선택이 가능하다. 상호억제의 특징은 뉴론이 서로 접속되어 있기 때문에 뉴론의 입력의 위치에 따라서 흥분기능과 억제기능을 가능하게 한다. 그림 3은 감각수용기에서의 결합신경의 하중을 나타내고 있으며, "+"는 흥분성세포에 의한 결합신경의 강도를 나타내며, "-"는 억제성세포에 의한 결합신경의 강도를 나타낸다. 그림 4는 그림 3의 흥분과 억제기능을 수용한 상호억제로써, 음성의 스펙트럼의 높은 부분을 강조하는 1개의 흥분영역과 낮은 부분의 잡음을 경감하는 2개의 억제영역을 가지며 주파수영역에서 사용한다. 그림 4에서 가로축은 주파수표본점을 나타내고, 세로축은 주파수 $B_f = 0$의 위치에 입력 1이 부가된 경우에 그 근방의 표본점에서 얻어진 출력을 나타낸다. 여기서 B_f는 상호억제폭을 결정하는 요소이며, $B_f = 5$인 경우이다.

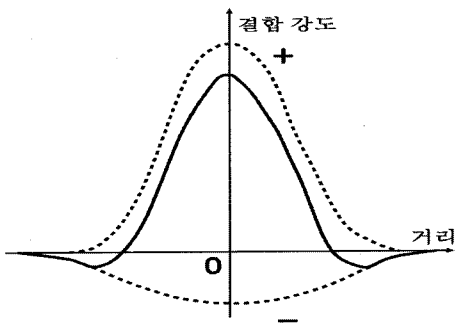

그림 3. 감각수용기에서의 결합신경의 하중
Fig. 3 Weight of connective neuron in sensory reception.

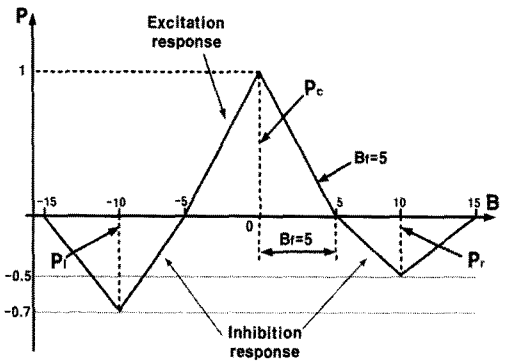

그림 4. 상호억제 함수의 임펄스 모델
Fig. 4 A functional representation of lateral inhibition.

그림 4에서 상호억제의 진폭을 나타내는 요소 P_l, P_c, P_r에 대하여 식 (1)의 제한을 설정한다.

$$P_l + P_c + P_r = 0 \quad \cdots\cdots (1)$$

식 (1)의 제한은 상호억제에 의해 잡음의 합의 평균값이 영으로 되어서 잡음이 억제된다.

Ⅳ. 제안한 적응적인 잡음억제 시스템

본 논문에 사용한 적응적인 잡음억제 시스템의 구성을 그림 5에 나타낸다.

먼저 잡음이 중첩된 음성신호는 한 프레임이 128 샘플로 구성되는 해밍창을 통과한 후 FFT되며 이 FFT된 신호는 유성부, 무성부 및 묵음부로 판별된다. 각 프레임에서, $R_f \geq T_h$일 때에는 이 프레임은 유성부로, $T_h < R_f \leq T_h/\alpha$일 때에는 이 프레임은 무성부로, $R_f < T_h/\alpha$일 때에는 이 프레임은 묵음부로 각각 판별된다. 여기에서 R_f는 각 프레임에서 구해진 실효값을 나타낸다. 본 실험에서는 처음의 약 5프레임에서 각 문장의 평균 실효값 R_m을 구하여 이 실효값이 문턱값 T_h가 되도록 실험적으로 정하였다. 또한 α는 3.0으로 하여 실험하였다. 각 프레임이 판별된 후에, 유성부에서는 직류성분을 포함하는 0번째부터 9번째까지의 10개의 켑스트럼 성분을 FFT한 후에 5프레임의 이동평균을 취하여 $B_f = 6$를 사용하여 상호억제를 한다. 무성부에서는 켑스트럼 변환 및 3프레임의 이동평균을 취한 후에 log power spectrum을 하며, 이 log 신호를 $B_f = 5$를 사용하여 상호억제한다. 묵음부에서는 5프레임의 이동평균을 한다. 본 실험에서는 식 (2)와 같은 가중치가 부가된 이동평균을 제안한다.

$$\overline{P}(i,\omega) = \frac{1}{2M+1}\sum_{j=-M}^{M} W_j P(i-j,\omega) \cdots\cdots (2)$$

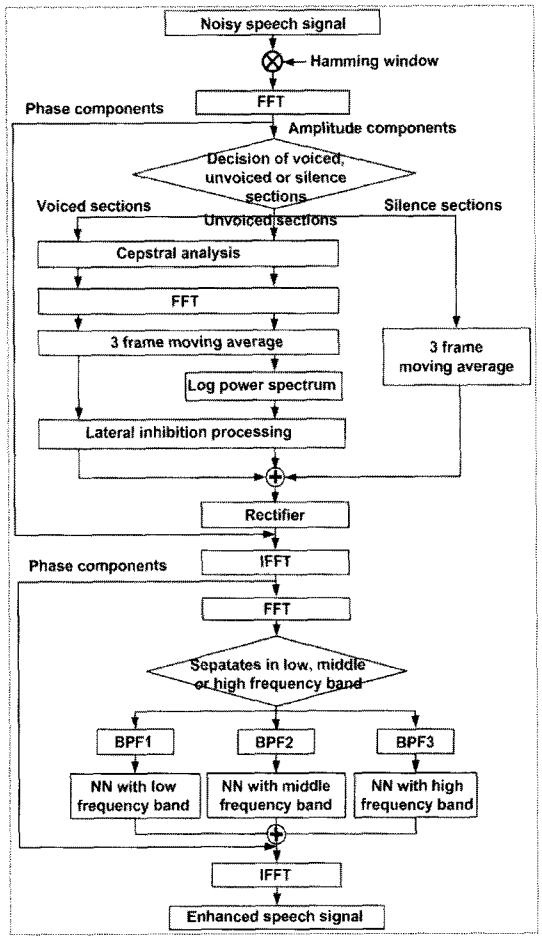

그림 5. 제안한 잡음억제 시스템
Fig. 5 The proposed noise suppression system.

본 실험에서는 $M=2$로 하며, 유성부와 무성부에서의 가중치는 $W_{-2}=0.6$, $W_{-1}=1.1$, $W_0=1.6$, $W_1=1.1$, $W_2=0.6$로 하고, 묵음부에서의 가중치는

$W_{-2}=0.7$, $W_{-1}=1.1$, $W_0=1.4$, $W_1=1.1$, $W_2=0.7$와 같이 각 구간에서 이동평균의 가중치를 다르게 적용하였다. 여기에서 $\overline{P}(i,\omega)$는 평균화된 (i)번째의 프레임의 단시간 전력스펙트럼이다. 이 후, 이 신호들을 합성하여 음의 성분만을 제거하는 정류기에 통과시킨 후에 IFFT한다. 이 신호를 다시 FFT하여 각각 저역, 중역, 고역의 신호로 분리한 후, 각 대역에서 진폭 및 위상성분으로 구성된 그림 2의 신경회로망들을 학습시켜 음성신호를 강조한다. 본 실험에서는 FFT에 의해 구해진 진폭성분 및 위상성분 중, 0~20샘플(0 kHz~1.2 kHz)은 저역(BPF1)부의 입력신호로, 21~41샘플(1.3 kHz~2.5 kHz)은 중역(BPF2)부의 입력신호로, 42~63샘플(2.6 kHz~3.9 kHz)은 고역(BPF3)부의 입력신호로 분할되어 입력되어 신경회로망에 의하여 학습된다. 따라서 각 대역은 약 2.63ms의 길이를 가진 샘플들로 구성되어 있다. NN의 입력신호에는 잡음이 중첩된 음성신호로부터 구해진 FFT 진폭 및 위상성분이 부여되며 학습신호에는 잡음을 부가하지 않은 음성신호로부터 구해진 FFT 성분을 부여하여 1프레임마다 학습을 한다.

V. 실험 및 결과

본장에서는 신경회로망 및 잡음억제시스템을 사용하여 음성 데이터에 대한 잡음제거의 실험결과에 대해서 기술한다. 본 실험에서는 시간영역의 평가척도인 SNR_{out}(Output SNR)을 사용하여 본 방법의 유효성을 확인한다. 본 시스템의 성능평가를 위하여, Aurora2 데이터베이스의 테스트셋 A, B, C로부터 잡음이 중첩된 음성데이터들이 임의적으로 선택되었다. 제안한 시스템은 정상잡음인 백색잡음(white noise) 및 자동차잡음(car noise), 그리고 비정상잡음인 도로잡음(street noise) 등에 대하여 NN에 의한 방법 및 MMSE-LSA(minimum mean-square error log-spectral amplitude)[10] 등과 비교되었다. 이 MMSE-LSA는 통계적으로 독립적인 가우시안 랜덤변수들을 사용하여 음성과 잡음의 스펙트럼 성분들을 유도한다. MMSE-LSA 방법을 실행할 때의 프레임 길이는 64샘플(8ms)이며, 각 프레임에서 해밍창이 사용되었다. 그림 6, 7은 백색잡음, 도로잡음에 대하여 다양한 잡음레벨들($Input\ SNR=20\,dB\sim 0\,dB$)을 사용하여, 20개의 문장에 대한 SNR_{out}의 평균값을 나타내었다. 그림 6의 백색잡음에 대하여, 잡음이 중첩된 음성신호(Original noisy speech)와 비교하였을 때, MMSE-LSA의 SNR_{out} 최대 개선값은 약 8dB, 위상성분의 NN을 사용하지 않은 경우(NN without phase component)의 SNR_{out} 최대 개선값은 약 9dB, 본 방법은 약 11.5dB 개선되었다. 그리고 그림 7의 도로잡음에 대해서도 유사한 경향이 보여지고 있으며, 잡음이 중첩된 음성신호와 비교하였을 때 본 방법은 최대 9.7dB이 개선되었다. 그림에는 나타나 있지 않지만 자동차잡음에 대해서도 같은 경향이 보여졌으며, 잡음이 중첩된 음성신호와 비교하였을 때 본 방법은 약 최대 10.5dB 개선되었다. 따라서 그림에 나타낸 것과 같이 제안한 시스템은 잡음레벨이 낮았을 때보다 잡음레벨이 높았을 때에 양호한 개선결과를 보였다.

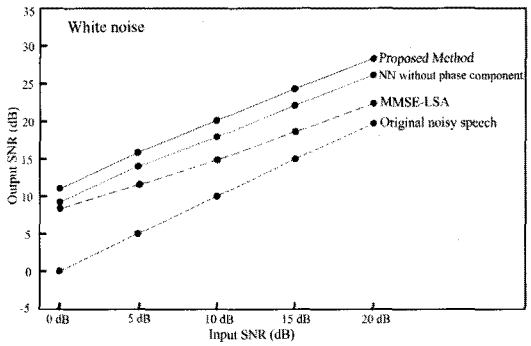

그림 6. 백색잡음 부가 시의 제안한 방식의 성능비교
Fig. 6 Comparison of the proposed method when adding white noise.

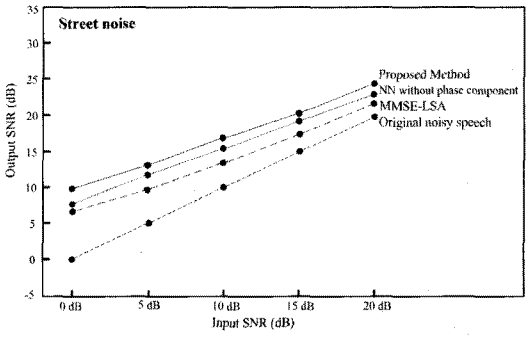

그림 7. 도로잡음 부가 시의 제안한 방식의 성능비교
Fig. 7 Comparison of the proposed method when adding street noise.

VI. 결론

배경잡음을 제거하기 위하여 적응적인 잡음억제시스템을 제안하여, 본 시스템이 백색잡음, 자동차잡음 및 도로잡음에 대해서 유효하다는 것을 SNR을 통하여 실험적으로 검증하였다. 따라서 제안한 시스템은 유성부, 무성부 및 묵음부에 대하여 각각 저역, 중역, 고역으로 분리된 신경회로망에 의하여 잡음이 제거됨을 확인할 수 있었다.

향후의 연구과제로는 신경회로망의 입력수가 많아짐에 따라 계산량이 증가하는 문제를 개선할 필요가 있으며, 입력샘플수를 증가시켰을 때에 학습능력을 향상시키기 위한 신경회로망의 학습조건을 변경시켜 학습시킬 필요가 있다고 본다. 이상으로, 본 논문에서 제안한 잡음에 강인한 잡음억제 시스템의 성과는 다양한 잡음하에서의 잡음억제 및 음성강조에 도움이 될 것으로 생각된다.

참고문헌

[1] N. Magotra, P. Kasthuri, Y. Yang, R. Whitman, F. Livingston, Multichannel adaptive noise reduction in digital hearing aids, Proc. of IEEE Int. Symp. Circuits Syst., Vol. 1998, No. 6, IV, pp. 582-585, 1988.

[2] B. Widrow, et al., "Adaptive noise cancelling: Principles and applications", Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, 1975.

[3] S. Tamura, "An analysis of a noise reduction neural network", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP vol. 89, no. 3, pp. 2001-2004, 1989.

[4] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise", IEEE Trans. Acoust., Speech, Signal Processing. Vol. 6, No. 5, pp. 471-472, 1978.

[5] 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, 1979.

[6] J. H. L. Hansen, 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, 1995.

[7] O. Ghitza, "Auditory neural feedback as a basis for speech processing", in Proc. Int. Conf. IEEE ASSP (New York, NY), pp. 91-94, 1988.

[8] Arslan, L., McCree, A. and Viswanathan, V., "New methods for adaptive noise suppression", IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-95), 812-815, 1995.

[9] Irino, T. and Patterson, R.D., "A time-domain, level-dependent auditory filter: The gammachirp", J. Acoust. Soc. Am. 101, 412-419, 1997.

[10] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 33, No. 2, pp. 443-445, 1985.

저자소개

최재승 (Jae Seung Choi)

1989년 조선대 전자공학과 공학사
1995년 일본 오사카시립대학 전자정보공학부 공학석사
1999년 일본 오사카시립대학 전자정보공학부 공학박사
2000년~2001년 일본 마쯔시타 전기산업주식회사 AVC사 연구원
2002년~2007 경북대 디지털기술연구소 책임연구원
2007년~현재 신라대학교 전자공학과 교수
※관심분야 : 디지털통신, 신호처리, 신경회로망, 적응필터와 잡음제거, 디지털 TV 및 멀티미디어 등