

잡음환경 하에서의 음성의 SNR 개선

최재승*

Improvement of Signal-to-Noise Ratio for Speech under Noisy Environment

Jae-Seung Choi*

Department of Electronic Engineering, Silla University, 140 Baegyong-daero 700beon-gil, Sasang-gu, Busan 617-736, Korea

요 약

본 논문에서는 잡음 환경 하에서 음성신호에 대한 신호대잡음비(SNR)를 개선하기 위한 알고리즘을 제안한다. 본 논문에서 제안하는 알고리즘은 백색잡음 및 자동차잡음 등과 같은 배경잡음으로부터 음성신호의 SNR을 개선할 목적으로 먼저 저역, 중역, 고역 SNR 대역에서 SNR을 추정한다. 다음으로 본 알고리즘은 각 대역에서 스펙트럼을 강조함으로써 잡음으로 오염된 음성신호 속에서 잡음신호를 차감한다. 백색잡음, 자동차잡음에 의하여 오염된 음성에 대하여 본 논문에서 제안한 알고리즘이 스펙트럼 차감 방법과 비교하여 양호한 신호대잡음비 값을 구하였다. 실험 결과로부터 스펙트럼 차감 방법과 비교하여 백색잡음에 대하여 최대 4.2 dB, 자동차잡음에 대하여 최대 3.7 dB의 출력 신호대잡음비가 개선된 것을 확인할 수 있었다.

ABSTRACT

This paper proposes an improvement algorithm of signal-to-noise ratios (SNRs) for speech signals under noisy environments. The proposed algorithm first estimates the SNRs in a low SNR, mid SNR and high SNR areas, in order to improve the SNRs in the speech signal from background noise, such as white noise and car noise. Thereafter, this algorithm subtracts the noise signal from the noisy speech signal at each bands using a spectrum sharpening method. In the experiment, good signal-to-noise ratios (SNR) are obtained for white noise and car noise compared with a conventional spectral subtraction method. From the experiment results, the maximal improvement in the output SNR results was approximately 4.2 dB and 3.7 dB better for white noise and car noise compared with the results of the spectral subtraction method, in the background noisy environment, respectively.

키워드 : SNR 개선 알고리즘, 신호대잡음비, SNR 추정 방법, 배경잡음

Key word : SNR improvement algorithm, signal-to-noise ratio, SNR estimation method, background noise

접수일자 : 2013. 03. 12 심사완료일자 : 2013. 04. 24 게재확정일자 : 2013. 05. 13

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

Department of Electronic Engineering, Silla University, 140 Baegyong-daero 700beon-gil, Sasang-gu, Busan 617-736, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2013.17.7.1571>

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. 서 론

음성신호는 잡음신호의 영향에 의하여 비선형적으로 변화한다. 이러한 변화로부터 원래의 음성신호로 복원하는 것이 가능하다면, 음성인식의 전처리로서 사용한다든가 잡음을 제거하는 것이 가능할 것으로 생각된다. 잡음으로 오염된 음성파형으로부터 직접 신경회로망 및 선형필터에 의하여 잡음을 제거하는 실험이 보고되어 있다[1, 2, 3]. 그리고 음성을 청취하는 경우, 실제 환경에 존재하는 잡음에 의해서 음성의 품질이 떨어지거나 음성의 명료도가 저감되어, 청취하는 사람이 자주 피로를 호소하는 경우가 있다. 이러한 문제를 해결하기 위하여 지금까지 많은 음성강조수법이 연구되어 왔다. 음성강조 기법은 시스템의 입력으로서 단일 채널의 음성을 사용하는 방법[4, 5], 다수 채널의 신호를 사용하는 방법[6, 7]의 두 가지로 크게 분류할 수 있다.

음성인식 등의 음성정보처리의 실용화에 대한 중요한 문제점으로 실제 환경에서의 잡음환경에 대한 대응이 생각되며, 여러 종류의 연구가 검토 진행되고 있다 [8, 9]. 이러한 대처방안으로는 입력파형을 직접적으로 처리하는 방법과 분석 파라미터를 처리하는 방법 등이 고려되고 있다. 전자로 생각되어지는 것은 복수의 마이크로폰 등을 이용하는 방법이 생각되어지고 있으며, 이 경우 적응 필터를 이용한 잡음제거 방법이 유효하며 여러 분야에서 검토되어지고 있다. 그러나 이 처리방식은 계산 등이 복잡하고, 잡음을 완전히 제거하는 것이 어렵다. 이러한 이유 때문에 이 방식을 음성처리 장치의 전처리로서 사용한다고 해도 여전히 이 외의 방법에 의한 잡음을 제거하는 처리방식이 필요하며, 이러한 방법으로는 주로 스펙트럼 영역에서 스펙트럼을 차감하는 방법 등이 있다.

본 논문에서는 잡음 환경 하에서의 음성신호의 잡음 제거에 응용하기 위해서 신호대잡음비(Signal-to-Noise Ratio, SNR)의 개선이 실현 가능하는 것을 목표로 한다. 본 논문에서는 잡음으로 오염된 음성신호의 입력을 음성신호와 잡음신호를 분리하여, 결과적으로 잡음을 저감하는 연구를 진행한다. 따라서 본 논문에서 제안하는 잡음저감 방법은 음성신호의 화자 및 잡음원에 대한 사전 정보는 필요하지 않으며, 비교적 높은 레벨에서의 비정상적인 잡음에 대해서도 유효하다는 것을 실험으로 나타낸다.

본 논문의 구성은, 2장에서 기존의 잡음차감법인 잡음 저감법의 개략에 대해서 설명한다. 3장에서는 제안하는 SNR 개선 알고리즘에 대해서 기술한다. 4장에서는 실험에 사용한 음성과 잡음의 데이터베이스 및 실험 조건에 대해서 설명한다. 5장에서는 실험결과를 기술하며, 마지막으로 6장에서는 결론을 맺는다.

II. 기존의 잡음 차감 방법

본 장에서는 본 논문에서 제안하는 알고리즘의 기초가 되는 기존의 스펙트럼 차감법(Spectral subtraction, SS)에 대하여 기술한다. 이 스펙트럼 차감법[4]은 목적신호의 단구간 진폭 스펙트럼의 추정법의 하나이며, 잡음이 중첩된 입력 음성신호의 스펙트럼으로부터 잡음신호의 스펙트럼을 차감하는 것으로부터 목적신호의 스펙트럼을 추정할 수 있다.

음성신호를 $s(t)$, 잡음을 $n(t)$ 이라 하면, 관측된 신호 $y(t)$ 은 다음식과 같이 나타낼 수 있다.

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

이 신호를 시간축 상에서 짧은 구간으로 분할하여, 시간축에 통과시켜 단구간 푸리에 변환(Fourier transform)을 실시함으로써 입력신호의 단구간 전력스펙트럼 $|Y(k)|^2$ 를 구한다. 여기에서, 음성신호 $s(t)$ 와 잡음 $n(t)$ 는 서로 상관이 없다고 가정하며, 또한 잡음이 정상적이라고 가정한다. 스펙트럼 차감 방법에서는 다음식에 의하여 목적신호의 단구간 스펙트럼의 추정값 $|\hat{S}(k)|^2$ 를 구한다.

$$|\hat{S}(k)|^2 = |Y(k)|^2 - E[|N(k)|^2] \quad (2)$$

식 (2)로부터 단구간 진폭 스펙트럼의 추정값 $|\hat{S}(k)|$ 를 구하여, 이것에 입력신호의 위상을 부가하여 역 푸리에변환을 수행함으로써 구하고자 하는 목적신호를 회복한다. 기존의 스펙트럼 차감법의 블록도를 그림 1에 나타낸다. 스펙트럼 차감 방법에서는 목적신호의 단구간 진폭 스펙트럼을 추정할 때에 잡음을 정상으로 가정하여, 잡음의 평균 진폭 스펙트럼을 사용한다. 그러나 가령 정상적이어도 잡음의 단구간 진폭 스펙트럼은

평균값의 주위에 분산하고 있어서, 이 분산 때문에 목적음의 단구간 진폭 스펙트럼의 추정 정확도가 떨어지기도 한다. 이러한 문제를 해결하기 위한 수단으로서 본 논문에서는 유성음 구간에 대하여 저역 SNR, 중역 SNR, 고역 SNR로 분류하여, 각 대역에서 잡음의 단구간 진폭 스펙트럼을 각각 추정함으로써 스펙트럼을 차감하는 알고리즘을 제안한다.

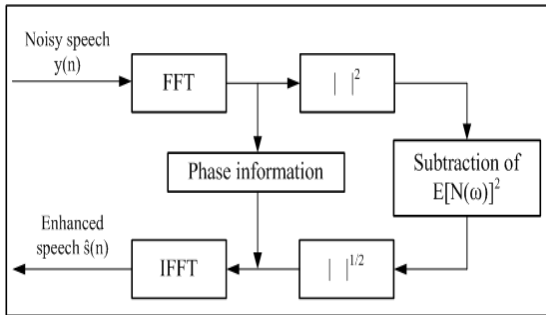


그림 1. 기존의 스펙트럼 차감법의 블록도
Fig. 1 Block diagram of conventional spectrum subtraction

III. 제안한 SNR 개선 알고리즘

최근, 잡음대책의 중요성이 주목되어 잡음과 음성을 구별하려는 시도의 연구가 계속적으로 수행되고 있다. 이러한 연구 수법으로는 주로 주파수영역에서의 스펙트럼의 차이를 중요시하는 경우가 일반적이다. 예를 들면, 신경회로망을 사용한 방법[2], 오염된 잡음의 스펙트럼을 차감하는 방법[4], 은닉 마르코프 모델(Hidden Markov Model, HMM)를 이용하는 방법[8], 그리고 캡스트럼 계수의 분포를 이용하는 방법[10] 등이 있다. 이러한 연구에 대해서는 주로 사무실 내에서 빈번하게 발생하는 여러 종류의 잡음을 대상으로 한 경우가 많으며, 다양한 잡음과 음성을 구별하기 위해서는 유효한 음성의 특징량 및 잡음차감 방법 등을 상세하게 검토할 필요가 있다. 따라서 본 논문에서는 잡음환경에 적절히 대응할 수 있도록 SNR 개선 알고리즘에 대하여 검토한다.

본 논문에서 제안한 SNR 개선 알고리즘의 블록도를 그림 2에 나타낸다.

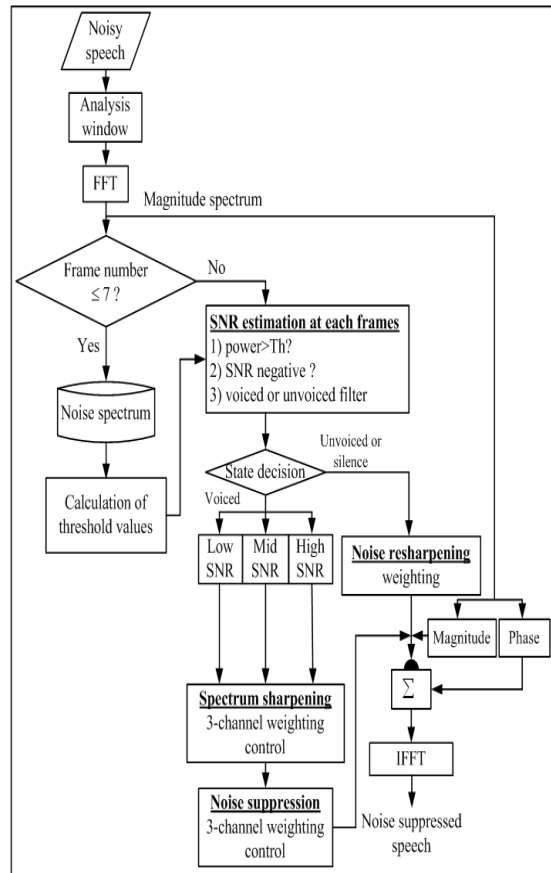


그림 2. 제안한 SNR 개선 알고리즘의 블록도
Fig. 2 Block diagram of proposed SNR improvement algorithm

먼저 잡음에 의해서 오염된 음성신호는 각 프레임으로 분리되어 해밍창에 통과시킨다. 이후에 고속 푸리에 변환(Fast Fourier Transform, FFT)을 하여 FFT의 진폭 성분에 대하여 각 프레임에서 전력 스펙트럼을 계산한다. 잡음으로 판단되는 처음의 7 프레임에서 계산된 잡음전력과 각 프레임의 전력 스펙트럼을 비교하여, 각 프레임의 전력 스펙트럼이 잡음전력보다 클 경우에는 음성신호로 판단하며, 반대의 경우에는 잡음신호로 판단한다. 그리고 SNR 추정 블록(SNR estimation at each frames)에서는 1) 전력 스펙트럼과 문턱값에 의한 비교 방법, 2) SNR의 부(음)의 성분 여부에 의한 비교 방법, 3) 유성음과 무성음의 필터에 의한 비교 방법 등을 사용하여 유성음 그리고 무성음(혹은 묵음) 구간을 결정한다. 여기에서 1)의 방법은 위에서 기술한 전력 스펙

트럼과 잡음전력 스펙트럼을 비교하는 방법과 동일하다. 2)의 방법은 각 프레임에서 SNR이 부(음)의 성분이면 무성음(혹은 묵음) 구간으로 우선 분류하며, 이와 반대이면 유성음 구간으로 분류한다. 3)의 방법에서는 저역 필터와 같은 형태의 필터로 설계한 것이 유성음 필터이며, 고역 필터와 같은 형태의 필터는 무성음 필터이다. 따라서 유성음 구간의 경우에는, 저역 SNR, 중역 SNR, 고역 SNR 대역으로 각각 분류하여, 각 대역별로 이동평균 및 잡음억제를 한다. 그리고 무성음(혹은 묵음) 구간의 경우에는 이동평균만을 실행한다. 이 후에 유성음 구간 및 무성음(혹은 묵음) 구간의 진폭성분을 합하며, 이때에 위상성분은 원래의 위상성분을 그대로 사용한다. 마지막으로 역 고속 푸리에 변환(Inverse Fast Fourier Transform, IFF)을 함으로써 잡음이 억제된 음성신호를 출력한다.

IV. 실험 조건 및 데이터베이스 모델

본 실험에서 사용한 음성신호는 8 kHz의 샘플링 주파수를 가진 환경에서 녹음된 영어숫자로 구성된 Aurora2 데이터베이스(Database, DB)[11]를 사용하였다. Aurora2 DB의 모든 음성데이터는 ETSI (European Telecommunications Standards Institute)로부터 배포되었으며, 남성화자 55명 및 여성화자 55명에 의해서 발성된 음성을 녹음한 총 8440개의 숫자로 구성된 테스트 셋 A, B, C의 음성데이터를 포함한다. 본 실험에서는 Aurora2 DB의 테스트 셋 A, B, C 중에서 임의적으로 남성화자 및 여성화자에 의한 10문장을 선택하여 평가용으로 사용하였다. 본 실험에 사용한 음성신호는 한 프레임의 분석구간의 길이가 16 ms이며, 분석구간의 데이터에 대하여 식 (3)의 해밍창을 통과시켰다. 여기에서 N은 분석구간의 음성데이터의 샘플수이다.

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (n = 0, \dots, N-1) \quad (3)$$

본 실험에서 사용한 잡음데이터는 Aurora2 DB의 자동차잡음(car noise)을 사용하였으며, 컴퓨터에 의해서 작성된 가우스 백색잡음(white noise)을 사용하여 평가하였다. 이러한 배경잡음은 샘플링 주파수 8 kHz이다.

잡음부가 음성은 위에서 기술한 음성데이터에 백색잡음 및 자동차잡음을 중첩하여 작성하였다. 각 음성데이터마다 서로 다른 백색잡음 및 자동차잡음을 중첩함으로써 입력 신호대잡음비(SNR_{input})가 20 dB, 15 dB, 10 dB, 5 dB의 잡음이 중첩된 음성을 작성하였다. 또한, SNR_{input}으로서는 다음 식에서 나타내는 바와 같이 음성 S(n)과 잡음 N(n)의 전체에 해당하는 전력의 비율로서 정의되는 전역 SNR_{input}을 사용하였다.

$$SNR_{input} = 10 \cdot \log_{10} \left(\frac{\sum_{n=1}^N S(n)^2}{\sum_{n=1}^N N(n)^2} \right) \quad (4)$$

본 논문에서 제안한 잡음제거의 평가방법으로는 식 (5)와 같은 출력 신호대잡음비(SNR_{output})를 사용하였다. 여기에서, $\hat{S}(n)$ 은 강조된 출력 음성신호이다.

$$SNR_{output} = 10 \cdot \log_{10} \left(\frac{\sum_{n=1}^N S(n)^2}{\sum_{n=1}^N [S(n) - \hat{S}(n)]^2} \right) \quad (5)$$

V. 실험 결과

본 논문에서 사용한 실험환경은 Intel(R) Core(TM)2 Quad 2.4GHz CPU와 3.25GB RAM이 장착된 IBM 호환 컴퓨터에서 Windows XP의 운영체제를 사용하여 Microsoft Visual C++ 6.0을 사용하여, C 언어로 프로그램을 구현하였다.

본 실험에서 평가용으로 사용하는 음성은 Aurora2 데이터베이스의 테스트 셋 A, B, C로부터 잡음이 중첩된 음성 데이터들이 임의적으로 선택되었으며, 잡음데이터는 컴퓨터에 의해서 작성된 가우스 백색잡음 및 Aurora2 데이터베이스의 테스트 자동차잡음(car noise)이 선택되었다.

제안한 알고리즘은 정상적인 잡음인 백색잡음 및 자동차잡음에 대하여 출력 SNR을 사용하여 기존의 SS 방법[4]과 비교하였다. 이 Boll에 의한 SS 방법은 2장의 기존의 SS 방법에서 자세히 기술한 바와 같이 잡음이

중첩된 입력 음성신호의 스펙트럼으로부터 잡음신호의 스펙트럼을 차감하는 방법으로 알려져 있다. 특히 Boll의 스펙트럼 차감법은 다소 최근에 제안된 방법은 아니지만 지금도 여러 분야에서 비교되어지는 방법 중의 하나이므로, 충분히 본 논문의 방법과 비교가 가능하다고 판단된다. 그러나 향후의 연구에서는 좀 더 최근에 발표된 기존 방법과의 비교가 필요하다고 본다.

표 1과 표 2는 백색잡음, 자동차잡음에 대하여 입력 SNR을 20 dB, 15 dB, 10 dB, 5 dB로 조정하였을 때, Aurora-2 데이터베이스에서 임의로 10개의 문장을 선택하여 각 잡음에 대하여 출력 SNR의 값을 계산한 평균 결과값을 나타내고 있다. 백색잡음이 음성신호에 중첩된 경우의 표 1의 결과를 보면, 본 논문에서 제안한 방법이 SS 방법보다 최대 4.2 dB 개선된 것을 알 수 있다. 자동차잡음이 음성신호에 중첩된 경우의 표 1의 결과에서는, 본 논문에서 제안한 방법이 SS 방법보다 최대 3.7 dB 개선된 것을 알 수 있다. 따라서 본 논문에서 제안한 SNR 개선 알고리즘에 의하여 백색잡음 및 자동차잡음에 대하여, 입력 SNR이 5 dB까지의 경우에 SNR이 향상된 것을 확인할 수 있었다.

표 1. 백색잡음인 경우의 출력 SNR 실험 결과

Table. 1 Output SNR results in the case of white noise

Input SNR[dB]	Output SNR [DB]		
	SS	Proposed	Impr.
20 dB	12.8	15.3	2.5
15 dB	13.4	16.5	3.1
10 dB	10.9	14.5	3.6
5 dB	9.1	13.3	4.2

표 2. 자동차잡음인 경우의 출력 SNR 실험 결과

Table. 2 Output SNR results when in the case of car noise

Input SNR[dB]	Output SNR [DB]		
	SS	Proposed	Impr.
20 dB	12.2	14.3	2.1
15 dB	11.9	14.7	2.8
10 dB	10.1	13.5	3.4
5 dB	8.6	12.3	3.7

VI. 결 론

본 논문에서는 잡음 환경 하에서의 음성의 잡음 저감 방법으로서 음성의 SNR 개선 알고리즘을 제안하였다. 백색잡음 및 자동차잡음에 의해서 오염된 음성신호에 대해서, 본 알고리즘을 사용하여 잡음제거 실험을 실시한 결과 신호대잡음비 레벨이 20 dB, 15 dB, 10 dB, 5 dB인 경우에 대하여 SNR 개선이 가능하였다. SS 방법과 비교하여 백색잡음에 대하여 최대 4.2 dB, 자동차잡음에 대하여 최대 3.7 dB의 출력 신호대잡음비가 개선된 것을 확인할 수 있었다. 그러나 본 논문에서의 SNR 개선 실험에서는, 음성데이터가 적은 것과 잡음원의 수가 2개로 제한된 것 등 충분한 실험조건은 아니었지만, 제안한 방법이 잡음 하에서의 음성의 SNR이 개선됨으로써 본 알고리즘이 유효한 것을 나타냈다. 향후에는 음성 데이터수를 증가시켜 실험을 실시하여, 보다 명확한 SNR 개선을 검토할 예정이다.

REFERENCES

- [1] S. Tamura and A. Waibel, "Noise reduction using connectionist models", *Proc. ICASSP-88*, Vol. 1, pp. 553-556, 1988.
- [2] T. T. Le, J. S. Mason and T. Kitamura, "Characteristics of multi-layer perceptron models in enhancing degraded speech", *Proc. ICSLP-94*, pp. 1611-1614, 1994.
- [3] J. S. Choi, "Speaker Recognition using LPC Cepstrum Coefficients and Neural Network", *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 15, No. 12, pp. 2521-2526, December 2011.
- [4] 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.
- [5] J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc., IEEE*, Vol. 67, DEC, 1979.
- [6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag*, AP-30, 27-34, 1982.
- [7] B. Widrow, R. John, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, R. C. Goodlin, "Adaptive noise cancelling:

- Principles and applications”, *Proc. IEEE*, Vol. 63, No. 12, pp. 1692-1716, 1975.
- [8] A. A. M. Abushariah, T. S. Gunawan, O. O. Khalifa and M. A. M. Abushariah, “English digits speech recognition system based on Hidden Markov Models”, *2010 International Conference on Computer and Communication Engineering*, pp. 1-5, May 2010.
- [9] L. Yang, L. Jing, Y. Yuxiang and W. Jian, “Improvement algorithm of DTW on isolated-word recognition”, *2011 IEEE International Conference on Computer Science and Automation Engineering*, Vol. 3, pp. 319-322, 2011.
- [10] J. S. Lim, “*Speech Enhancement*”, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1983.
- [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, 2000.



최재승(Jae-Seung Choi)

1989년 조선대학교 전자공학과 공학사
1995년 일본 오사카시립대학교 전자정보공학부 공학석사
1999년 일본 오사카시립대학교 전자정보공학부 공학박사
2000년 ~ 2001년 일본 마쯔시타(현, 파나소닉 주식회사) 전기산업주식회사 AVC사 전임연구원
2002년 ~ 2007 경북대 디지털기술연구소 책임연구원
2007년 ~ 현재 신라대학교 전자공학과 교수
※관심분야 : 음성신호처리, 적응신호처리, 잡음제거, 신경회로망, 디지털통신, 멀티미디어 등