인간의 청각모델에 기초한 잡음환경에 적응된 잡음억압 시스템

최재승*

*신라대학교

Adaptive Noise Suppression system based on Human Auditory Model

Jae-seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 다양한 배경잡음에 의해 열화된 음성을 강조하기 위하여 청각모델에 기초로 한 잡음환경에 적응된 잡음억압 시스템을 제안한다. 제안한 시스템은 먼저 유성음과 무성음의 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 and unvoiced sections for each frame and implements the adaptive auditory process, then reduces the noise speech signal using neural network including amplitude component and phase component. Base 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, neural network, speech enhancement

Ⅰ. 서 론

근대과학의 진보에 따라서 컴퓨터 관련 기술의 진보에 의하여 정보처리에 관한 분야가 눈부시게 발전하고 있다. 이러한 분야 중에서 신호처리, 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 잡음억압시스템은 음성 혹은 잡음의 성질에 기초를 두었다. 적응적인 잡음 제거(Adaptive noise canceling)[1, 2]는 음성 혹은 잡음에 상관이 있는 적응적으로 가중치를 적용한 참조신호에 의해서, 최소 2승 평균을 사용하여 잡음을 감소시켜 음성신호를 강조한다. Adaptive comb filtering[3]은 잡음으로 추정되어지는 비주기성분을 제거하기 위하여 유성음의 주기성을 이용한다. 적당한 데이터베이스에 기초한 강조함수를 학습하는 데는 신경회로망을 사용한다[4]. Spectral subtraction[5, 6]은 음성신호가 잡음과 무상관이라고 가정하므로 잡음을 포함한 신호의 전력스펙트럼은 음성과 잡음의 스펙트럼의 합이다. 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.

수많은 연구자에 의한 수년간의 노력에도 불구하고, 잡음제거 및 음성강조의 연구는 아직 완성되어 있지 않다고 볼 수 있다. 이것은 다음과 같은 문제를 동시에 해결하려고 한 것이기 때문이다. 1) 잡음레벨을 추정하기 위해서 잡음을 포함한 음성의 비음성 구간의 검출을 필요로 한다. 2) 무성음을 강조하는데에 대한 실패, 3) 비정상적인 방해잡음을 감소시키는데에 대한 실패, 4) 계산량의 많음, 5) 시스템의 학습조건에의 의존성 등을 들 수 있다. 본 논문에서는, 상기 문제의 2), 3), 5)를 해결하여, 남은 문제의 효과를 최대한 줄이기 위한 알고리즘을 제공한다. 알려진 바와 같이, 인간의 청각시스템은 배경잡음을 압축하는 것이 가능하고, 음성과 잡음에 대한 사전의 지식없이 희망하는 신호를 선택이 가능하다. 귀의 강조기능을 시뮬레이트하는 것에 의해서 음성의 강조를 시도하는 것은 당연히 자연스러운 것이다. Ghiza[7]는 완전한 와우각(달팽이관)모델을 사용하여, 음성신호는 청각의 처리를 통하여 강조되는 것을 나타내고 있다. 그러나 와우각 모델은 다수의 청각기능을 가지고 있어서, 이것들의 기능의 모두가 음성강조에 유익하지는 않은 것 같다. 더욱이 이와 같은 귀의 모델은 계산량이 방대하다. 본 연구는 참고문헌[8]에서 사용한 상호억제(lateral inhibition)라고 불리우는 하나의 청각모델을 사용하여 연구한다. 이것은 생리학 및 심리학을 통해서 발견되어, 신경생리학은 이것을 음성의 스펙트럼을 날카롭게 하는 것으로부터 관계되었다.

본 논문에서는 잡음이 존재하는 환경 하에서 먼저 상호억제라고 하는 청각기강을 공학적으로 응용하는 방법을 제안하며, 시간지연 신경회로망(Time Delay Neural Network: TDNN)[9, 10]에 시간요소뿐만 아니라 위상요소도 도입한 TDNN을 도입하여 고속 푸리에 변환(fast Fourier transform : FFT)한 진폭성분 및 위상성분을 복원하는 알고리즘을 제안한다. 그리고 마지막으로 유성음과 무성음의 구간을 검출한 후, 저역, 중역, 고역으로 분리된 신경회로망을 사용하여 잡음신호를 제거한 후에 음성을 강조하는 처리를 한다.

II. 본 론

계산기에 의한 청각특성의 분석에 있어서, 음성 및 음악의 배경잡음 등의 정상적인 배경잡음을 억압하는 것은, 목적으로 하는 신호를 고정도로 추출하기 위해서도 중요하다. 이와 같은 목적으로 사용가능한 잡음억압처리로 가장 먼저 인용되어지는 방식은, Boll에 의한 스펙트럼 차감법[6]이다. 그러나 이 스펙트럼 차감법은 청취자를 상정한 분석 합성계를 사용하는 경우에는 그다지 사용하고 있지 않다. 이것은 잘 알려진 바와 같이, 합성음에 "musical noise"(악음적 잡음)이 발생하기 때문에, 강조하였다고 생각한 목적 신호가 도리어 듣기 어려워지기 때문이다. 이러한 것은 악음적 잡음과 같은 곳에서 일어나던가, 예측할 수 없는 잡음이 있으면 인간의 우수한 청각계의 기능을 도리어 방해하는 것을 말한다.

이 악음적 잡음의 결점을 극복하기 위해서는 몇 가지 새로운 개량방법이 제안되어 있지만[11, 12, 13], 신호대잡음비(Signal-to-Noise Ration; SNR)이 0 dB에 가까운 경우의 SNR 개선도의 평가에 대한 결과는 없으며, 항상 유효한 방법이 되는가에 대해서도 정확히 알 수 없다. 원래 스펙트럼 차감법에 대해서는, 스펙트럼의 진폭성분과 위상성분을 분리하여, 진폭성분만을 조작하며 위상성분은 그대로 사용하고 있다. 이 처리에 의한 진폭성분과 위상성분의 물리적인 부정합이 악음적 잡음의 원인이 되며, 이것을 해소하지 않는 한 본질적인 해결을 할 수 없다.

본 논문에서는 이 문제점을 본질적으로 해결하는 잡음에 적응적인 잡음억압법을 제안한다. 이 방법은 스펙트럼 차감법과 동일한 전제 조건으로 동등의 SNR 개선이 가능하며, 그러나 악음적 잡음을 발생시키지 않는 방법이다. 따라서 본 논문에서는 잡음억압을 위한 시불변의 단시간 푸리에 변환뿐만 아니라 시간 및 위상성분이 도입된 그림 1의 신경회로망 시스템을 제안한다.

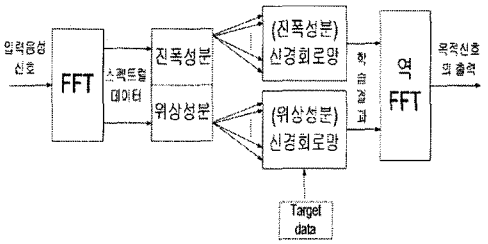

그림 1 제안한 신경회로망 시스템

그림 1은 본 논문에서 제안한 진폭성분 및 위상성분으로 분리된 입력층, 중간층, 출력층을 가진 각 유닛수가 64개인 3층의 신경회로망 시스템을 나타낸다. 본 시스템은 64샘플의 입력 데이터에 대해서 FFT를 실시하여 스펙트럴의 데이터를 구한다. 이 스펙트럴 데이터를 진폭성분과 위상성분으로 분리하여 각 64샘플의 데이터를 진폭 및 위상성분의 신경회로망의 입력으로 한다. 스펙트럴의 데이터를 각각 64입력, 64출력의 3층의 신경회로망에 입력함으로써, 각 출력신호는 학습신호와 일치한 정확한 값을 취하도록 학습한다.

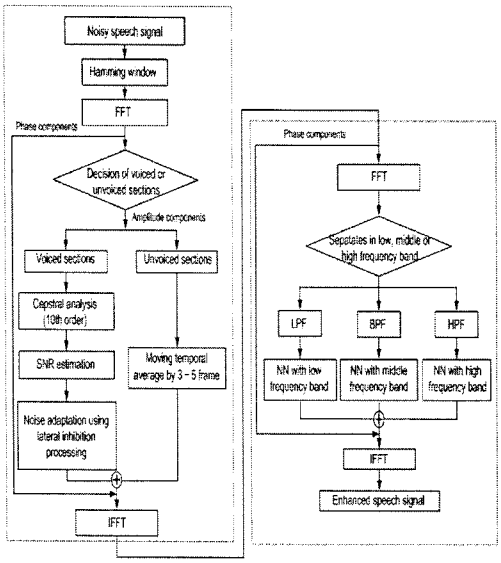

그림 2 제안한 적응적 잡음억압 시스템

본 논문에 사용한 적응적인 잡음억압 시스템의 구성을 그림 2에 나타낸다. 먼저 잡음이 중첩된 음성신호는 한 프레임이 64샘플로 구성되는 해밍창을 통과한 후 FFT되며 이 FFT된 신호는 유성음 및 무성음으로 판별된다. FFT된 신호의 유성음 구간에서는 직류성분을 포함하는 0번째부터 9번째까지의 10개의 캡스트럼 성분을 취한 후, SNR 추정에 의한 상호억제에 의한 적응적인 잡음억압 처리를 실시한다. 또한 FFT된 신호의 무성음 구간에서는 3프레임에서 5프레임에 해당하는 이동평균을 취한다. 따라서 이 두 신호를 합성하여 역 고속 푸리에 변환(Inverse Fast Fourier Transform: IFFT)한다. 이 신호를 다시 FFT하여 각각 저역, 중역, 고역의 신호로 분리한 후, 각 대역에서 진폭성분 및 위상성분으로 구성된 그림 1의 신경회로망을 사용하여 잡음이 억압된 신호를 출력할 수 있도록 신경회로망들을 학습시켜 음성신호를 강조한다.

III. 실험 및 결과

본 논문은 신경회로망 및 잡음억압 시스템을 사용하여 음성신호에 중첩된 잡음을 제거하는 것을 목적으로 하여, 음성 데이터에 대한 잡음제거의 실험결과에 대해서 기술한다. 본 실험에서는 시간영역의 평가척도인 식 (1)의 신호대잡음비(Segmental Signal-to-Noise Ratio; SNR)을 사용하여 본 방법의 유효성을 확인한다.

$$SNR = 10log_{10}\frac{\sum_{i=1}^{N} s^2(i)}{\sum_{i=1}^{N}\{s(i)-\hat{s}(i)\}^2}(dB) \quad (1)$$

여기에서, $s(i)$, $\hat{s}(i)$는 각각 입력신호 및 출력신호의 표본값이며 N은 측정구간의 표본수(N=64)를 나타낸다.

본 실험에서는 Aurora2 데이터베이스를 사용하여 여러 잡음환경 하에서 제안한 시스템에 대한 성능평가를 나타낸다. 본 시스템의 성능을 평가하기 위하여, Aurora2 데이터베이스의 테스트셋 A, B로부터 잡음이 중첩된 음성데이터들이 임의적으로 선택되었다. 제안한 시스템은 백색잡음, 자동차잡음 등에 대하여 신경회로망에 의한 방법 등과 비교되었으며, 실행할 때의 프레임길이는 64샘플(8ms)로 하여 각 프레임에서 해밍창이 사용되었다.

그림 3과 4는 백색잡음과 자동차잡음에 대하여 다양한 잡음레벨들($Input\,SNR = 20\,dB \sim 0dB$)을 사용하여, 20개의 문장에 대한 SNR의 평균값을 나타내었다. 그림 3의 백색잡음에 대하여, 잡음이 중첩된 음성신호(Original noisy speech)와 비교하였을 때, 위상성분의 NN을 사용하지 않은 경우(NN without phase component)의 SNR의 최대 개선값은 약 8 dB, 본 방법은 약 11 dB 개선되었다. 그리고 그림 4의 자동차잡음에 대해서도 같은 경향이 보여져, 잡음이 중첩된 음성신호(Original noisy speech)와 비교하였을 때, 위상성분의 NN을 사용하지 않은 경우(NN without phase component)의 SNR의 최대 개선값은 약 7 dB, 본 방법은 약 9.5 dB 개선되었다. 이상의 결과로부터, 잡음이 중첩된 입력음성신호에 대해서도 본 시스템이 충분히 SNR을 개선함으로써 본 시스템의 성능개선을 확인할 수 있었으며, 여러 잡음에 대하여 유효하다는 것을 말할 수 있다.

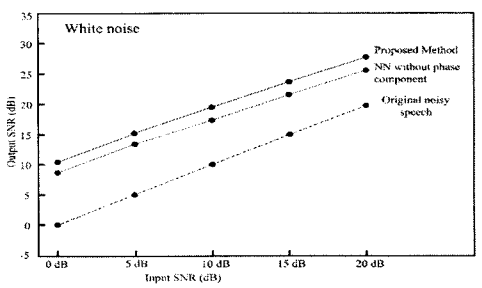

그림 3. 백색잡음 부가 시의 제안한 시스템과 NN과의 비교

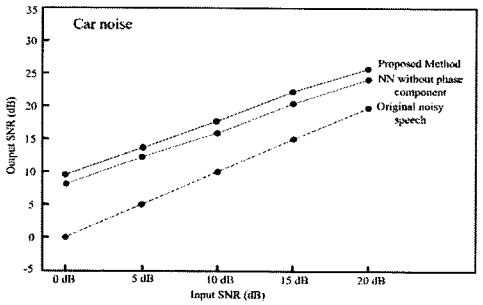

그림 4. 자동차잡음 부가 시의 제안한 시스템과 NN과의 비교

IV. 결론

배경잡음을 제거하기 위하여 적응적 음성강조시스템을 제안하여, 본 시스템이 백색잡음 및 자동차잡음에 대해서 유효하다는 것을 SNR을 사용하여 실험적으로 검증하였다. 따라서 제안한 적응적인 음성강조시스템은 유성음 및 무성음에 대하여 각각 저역, 중역, 고역부로 분리된 신경회로망에 의하여 잡음이 제거됨을 확인할 수 있었다.

향후의 연구과제로서는 각 신경회로망에 입력되어지는 입력수가 많아짐에 따라 계산량이 증가하는 문제를 개선할 필요가 있으며, 입력 샘플수를 증가시켜 신경회로망의 학습 조건을 변경시켜 학습능력을 향상시키는 시도가 필요하다고 본다. 그리고 본 논문에서는 위상성분에 해당하는 신경회로망도 사용하였지만 좀 더 개선된 신경회로망을 구성하고자 한다.

이상으로, 본 논문에서 제안한 잡음에 강인한 잡음억압 시스템의 성과는 다양한 잡음 하에서의 잡음억압 및 음성강조에 도움이 될 것으로 생각된다.

참고문헌

[1] 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.

[2] M. R. Sambur, "Adaptive noise cancelling for speech signals", IEEE Trans. Acoust., Speech, Signal Processing. Vol. 26, No. 5, pp. 419-423, 1978.

[3] J. S. Lim, A. V. Oppenheim, L. D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition", IEEE Trans. Acoust., Speech, Signal Processing, vol. 26, no. 4, pp. 354-358, 1978.

[4] 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.

[5] 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.

[6] 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.

[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] 최재승, "상호억제와 시간지연 신경회로망을 사용한 적응적인 음성강조시스템", 대한전자공학회 논문지, 제45권 2호 SP편, pp. 95-102, 2008. 3.

[9] M. Miyatake, H. Sawai, and K. Shikano, "Training Methods and Their Effects for Spoting Japanese Phenomes Using Time-Delay Neural Networks", IEICE, Vol. J73-D-II, No.5, pp. 699-706, 1990.

[10] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, Phoneme Recognition using Time-delay Neural Networks, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 3, pp. 328-339, 1989.

[11] 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.

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

[13] Irino, T. and Unoki, M., "A time-varying, Analysis/synthesis auditory filterbank using the gammachirp," IEEE Int. Conf. Acoust., Speech Signal Processing (ICASSP-98), 1998.