잡음억제 신경회로망에 의한 스펙트럼의 추정 기법

최재승*

*신라대학교 전자공학과

E-mail : *jschoi@silla.ac.kr

요 약

음성인식 및 음성신호처리 분야에서 신경회로망은 음성인식의 카테고리 분류에 주로 이용되고 있다는 점에 착안하여, 본 논문에서는 신경회로망의 입력신호로 음성의 진폭 스펙트럼 및 위상 스펙트럼을 사용한 잡음억제를 위한 신경회로망을 제안한다. 본 논문에서 제안한 알고리즘은 고속 푸리에 변환(Fast Fourier Transform, FFT)에 의한 진폭 스펙트럼 및 위상 스펙트럼을 사용한 잡음억제 신경회로망을 이용하여 각 프레임에서 FFT 스펙트럼을 추정한다.

키워드

음성처리, 신경회로망, 스펙트럼, 잡음억제

Ⅰ. 서 론

근년, 실제 환경에 존재하는 다양한 배경잡음이 사람이 발성하는 음성의 중요한 특징량의 왜곡은 물론 음성 구간의 추출 정밀도를 열화시킴으로써 음성 인식률을 저하시켜 음성인식 시스템의 성능을 나쁘게 하고 있다[1, 2]. 이와 같이 다양한 배경잡음이 존재하는 실제 환경에서 음성인식을 구현하기 위해서는 지금까지의 음성인식 알고리즘에 대한 잡음성능의 강건성 향상이 필요하게 되고 있으며, 이에 따라서 근년 이러한 주제의 연구가 지속적으로 증가하고 있는 추세이다[3, 4, 5, 6].

음성신호처리 분야에서 신경회로망[7, 8]은 음성인식의 카테고리 분류에 주로 이용되고 있다. 이러한 음성인식의 카테고리 분류에 이용되는 신경회로망의 입력신호로는 음성의 진폭 스펙트럼이 주로 많이 사용되고 있다. 그러나 본 논문에서는 진폭 스펙트럼뿐만 아니라 위상 스펙트럼의 중요성도 이용하여 잡음억제를 위한 알고리즘을 제안한다. 본 논문에서 제안한 알고리즘은 고속 푸리에 변환(Fast Fourier Transform, FFT)에 의한 진폭 스펙트럼 및 위상 스펙트럼을 사용한 잡음억제 신경회로망을 이용하여 각 프레임에서 FFT의 진폭 및 위상 스펙트럼을 추정한다.

Ⅱ. 제안한 잡음억제 신경회로망

본 논문에서 제안한 그림 1의 잡음억제 신경회로망(Noise Reduction Neural Network, NRNN)은 FFT 진폭 및 위상 스펙트럼 영역으로 구성된다. FFT의 진폭 및 위상 스펙트럼을 추정하기 위하여, 본 실험에서는 오차 역전파 학습알고리즘을 사용한 그림 2와 같은 퍼셉트론형의 잡음억제 신경회로망을 사용하여 학습시킨다. 신경회로망의 구성은 3층으로 구성되며, 32-64-32의 네트워크를 사용한다. 신경회로망의 학습 시에, 학습계수는 0.1, 관성계수는 0.03으로 하며, 최대 반복학습 횟수는 10,000회로 하였다. 본 실험에서는 FFT 진폭 스펙트럼 및 위상 스펙트럼에 대한 학습시의 신경회로망의 입력 신호는 0부터 31샘플(0 kHz부터 3.9 kHz)이며, 타겟 신

호는 음성신호(Clean signal)에 대하여 학습신호와 동일하게 0부터 31샘플(0 kHz부터 3.9 kHz)이다.

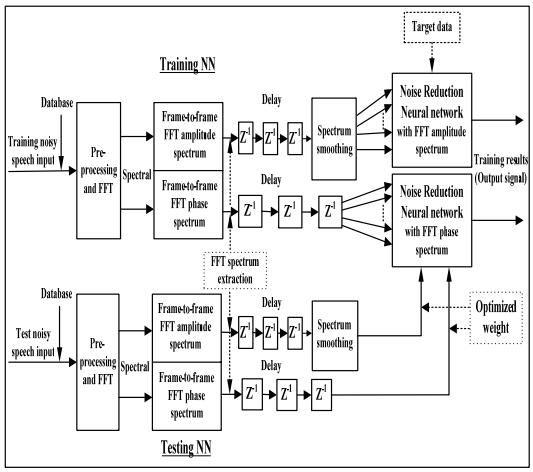

그림 1. 제안한 잡음억제 신경회로망(NRNN) 시스템

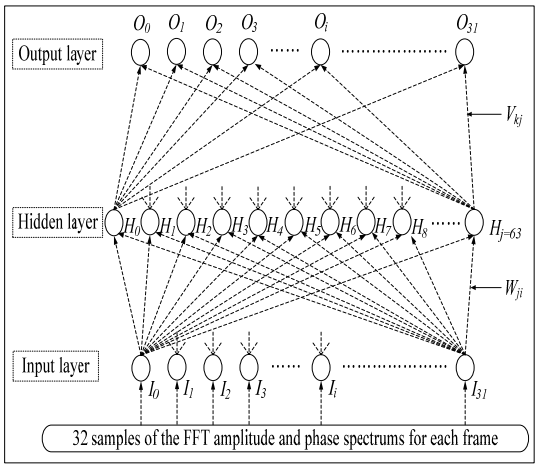

그림 2. 3층의 신경회로망의 구조

III. 실험 결과

잡음억제 신경회로망의 성능을 나타내기 위하여, Aurora-2 DB의 테스트 셋 A로부터 랜덤하게 선택된 남성화자의 음성신호 "MAT_32O2A.08"에 자동차잡음을 중첩한 10 프레임의 그래프를 그림 3과 그림 4에 나타낸다. 그림 3은 SNR=5 dB에 대하여 타겟 신호에 대한 FFT 진폭 스펙트럼(clean speech), 입력신호(noisy speech), 출력신호를 나타내며, 그림 4는 타겟 신호에 대한 FFT 위상 스펙트럼(clean speech), 입력신호(noisy speech), 출력신호를 각각 나타낸다. 그림에서 알 수 있듯이 제안한 NRNN 알고리즘에 의하여 FFT의 진폭 및 위상 스펙트럼을 추정한 것을 알 수 있다.

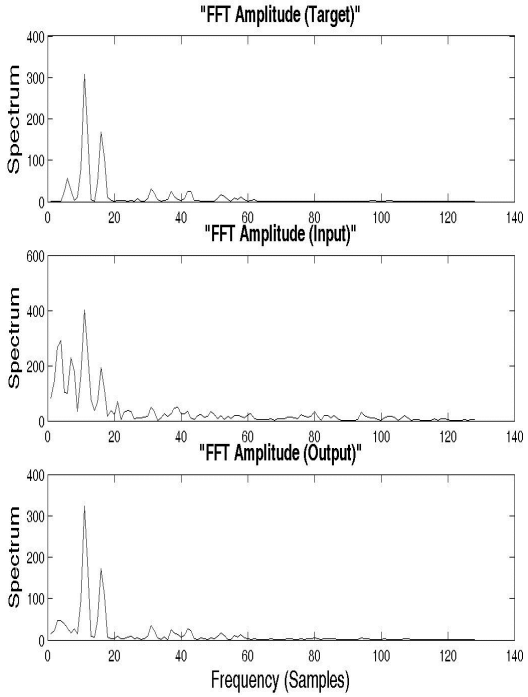

그림 3. 타겟, 입력, 출력신호에 대한 FFT 진폭 스펙트럼의 비교

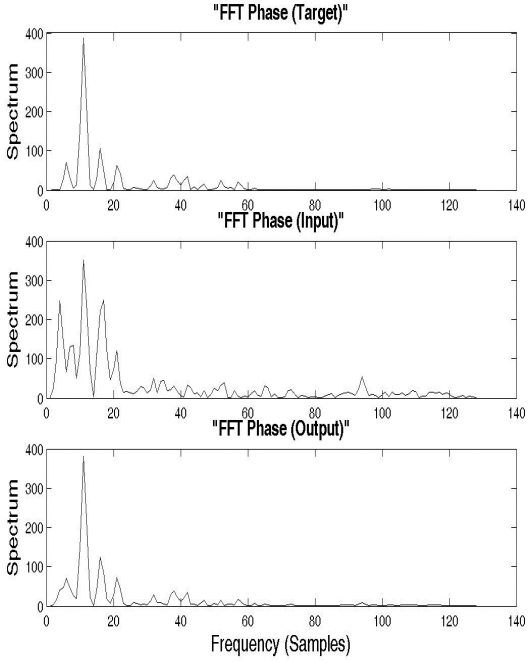

그림 4. 타겟, 입력, 출력신호에 대한 FFT 위상 스펙트럼의 비교

Ⅳ. 결론

본 논문에서는 배경잡음을 억제하기 위하여, FFT 진폭 및 위상 스펙트럼의 요소로 구성되는 잡음억제 신경회로망을 제안하였다. 본 실험에서는 오차 역전파 학습 알고리즘을 사용한 퍼셉트론형의 잡음억제 신경회로망을 사용하여 학습시켰다. 신경회로망의 구성은 3층으로 구성된 32-64-32의 네트워크를 사용하였다. 따라서 본 논문에서 제안한 알고리즘에 의하여 고속푸리에 변환에 의한 진폭 스펙트럼 및 위상 스펙트럼을 사용한 잡음억제 신경회로망을 이용하여 각 프레임에서 FFT의 진폭 및 위상 스펙트럼을 추정할 수 있었다. 특히 자동차잡음에 대하여 본 알고리즘이 유효한 것을 실험적으로 확인하였다.

참고문헌

[1] J. P. Haton, "Automatic recognition of noisy speech," In A.J.R. Ayuso and J.M.L. Soler, Eds., Speech Recognition and Coding-New Advances and Trends, Springer Verlag, Berlin, Germany, pp.3-13, 1995.

[2] D. Yu, L. Deng, J. Droppo, J. Wu, Y. Gong and A. Acero, “A minimum-mean-square-error noise reduction algorithm on mel-frequency cepstra for robust speech recognition,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4041-4044, 2008.

[3] Simpson, et. al., "Spectral Enhancement to Improve the Intelligibility of Speech in Noise for Hearing Impaired Listeners," Acta Otolaryngol, Suppl. 469, pp. 101-107, 1990.

[4] Y. Wu and Y. Li, “Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method,” IEEE International Conference on Systems, Man, and Cybernetics, Vol. 4, pp. 2935-2939, 2000.

[5] 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.

[6] R. Okamoto, Y. Takahashi, H. Saruwatari and K. Shikano, "MMSE STSA estimator with nonstationary noise estimation based on ICA for high-quality speech enhancement," IEEE International Conference on Acoustics Speech and Signal Processing, pp. 4778-4781, 2010.

[7] K. Daqrouq, I. N. Abu-Isbeih and M. Alfauri, "Speech signal enhancement using neural network and wavelet transform," 6th International Multi-Conference on Systems, Signals and Devices, pp. 1-6, 2009.

[8] W. G. Knecht, M. E. Schenkel and G. S. Moschytz, “Neural network filters for speech enhancement,” IEEE Trans. Speech and Audio Processing, vol. 3, pp. 433-438, 1995.