신경회로망에 의한 음성스펙트럼의 복원 알고리즘

최재승*

*신라대학교 전자공학과

Restoration Algorithm of Speech Spectrum using Neural Network

Jae-Seung Choi*

*Department of Electronic Engineering, Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 스펙트럼 회복의 수단으로써 신경회로망을 사용하여 푸리에변환(FFT) 진폭성분 및 위상성분을 복원하는 알고리즘을 제안한다. 본 논문에서는 먼저 각 프레임의 FFT 진폭성분들을 유성음 구간과 무성음 구간으로 검출한 후, 유성음 및 무성음 구간에 대해서 각 프레임의 FFT 진폭성분들을 저역, 중역 및 고역으로 각각 분리한 후에 각 대역의 FFT 진폭성분들을 저역용 신경회로망(NN), 중역용 NN, 그리고 고역용 NN의 입력으로 하여 각 NN에 학습시킴으로써 최종 FFT 진폭성분들을 구한다. 본 실험에서는 Aurora2 데이터베이스를 사용하여 FFT의 진폭성분을 복원하는 잡음제거의 알고리즘을 사용하여 여러 잡음에 대해서 본 알고리즘의 유효성을 실험적으로 확인한다.

키워드

Neural network, Recovery algorithm, Speech spectrum, Speech signal.

Ⅰ. 서 론

음성인식 및 스펙트럼 회복을 목적으로 한 논문 중에서 적당한 데이터베이스에 기초한 강조함수를 학습하는 데는 신경회로망(Neural Network: NN)을 사용한다[1, 2]. Spectral subtraction[3, 4]에서는 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.

본 논문에서는 스펙트럼 회복의 수단으로써 신경회로망(Neural network)을 사용하여 푸리에변환(fast Fourier transform : FFT) 진폭성분 및 위상성분을 복원하는 알고리즘을 제안한다. 본 논문에서는 먼저 각 프레임의 FFT 진폭성분들을 유성음 구간과 무성음 구간으로 검출한 후, 유성음 및 무성음 구간에 대해서 각 프레임의 FFT 진폭성분들을 저역, 중역 및 고역으로 각각 분리한 후에 각 대역의 FFT 진폭성분들을 저역용 NN, 중역용 NN, 그리고 고역용 NN의 입력으로 하여 각 NN에 학습시킴으로써 최종 FFT 진폭성분들을 구한다. 본 실험에서는 Aurora2 데이터베이스를 사용하여 FFT의 진폭성분을 복원하는 잡음제거의 알고리즘을 사용하여 여러 잡음에 대해서 본 알고리즘의 유효성을 실험적으로 확인한다.

Ⅱ. 음성스펙트럼의 회복 알고리즘

제안한 시스템은 영어숫자로 구성된 Aurora2 데이터베이스(테스트 셋 A, B, C 포함)[5]로부터의 테스트 셋 A와 B의 음성데이터와 컴퓨터에 의해서 생성한 가우스 백색잡음을 사용하여 평가하였다. 본 실험에서 다양한 신호대잡음비

(Signal-to-Noise Ratio: SNR=20dB, 15dB, 10dB, 5dB)이 부가된 잡음이 중첩된 음성신호를 사용하여 NN이 학습되었다. 본 실험에서의 음성 및 잡음 데이터베이스는 8 kHz의 표본 주파수를 가진다.

본 논문에서는 유성음 및 무성음 구간에 대하여 저역부, 중역부 및 고역부에 해당하는 주파수 대역별 NN의 알고리즘을 그림 1과 같이 제안하며 신경회로망은 오차 역전파알고리즘[6]을 사용하여 학습한다.

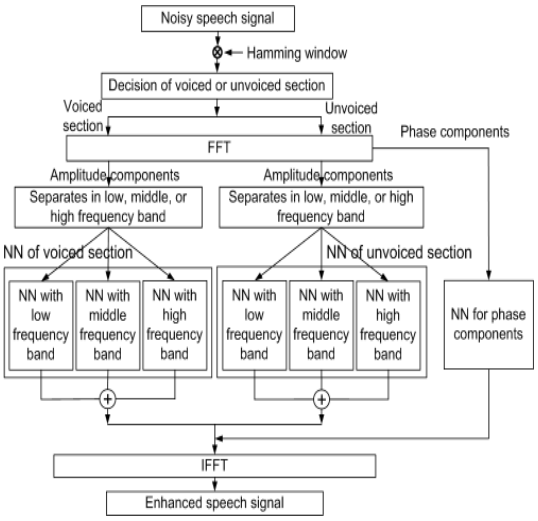

그림 1 제안한 주파수 대역별 NN 시스템

먼저 배경잡음이 중첩된 음성신호는 한 프레임이 128 샘플로 구성되며 각 프레임에 대하여 해밍창(hamming window)를 곱한다. 각 프레임에서 유성음 및 무성음을 검출한 후에, 각 유성음 및 무성음 구간에 대하여 FFT의 진폭성분들이 저역, 중역, 고역으로 분리되며 분리된 FFT 진폭성분들은 각 대역의 NN의 입력으로 부가되어 그림 1의 NN에 의하여 학습을 한다. 각각의 저역용 NN, 중역용 NN 및 고역용 NN으로부터의 출력을 합성하여 최종 FFT 진폭성분을 구한다. 그러나 위상성분은 별도의 NN에 의하여 학습되어 위상성분을 구한다. 마지막으로 역 고속 푸리에 변환(inverse fast Fourier transform : IFFT)을 사용하여 강조된 음성신호를 구한다.

본 실험에서는 FFT에 의해 구해진 진폭성분 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 NN에 의하여 학습된다. NN의 입력신호에는 잡음이 중첩된 음성신호로부터 구해진 FFT 진폭성분이 부여되며 학습신호에는 잡음을 부가하지 않은 음성신호로부터 구해진 FFT 성분을 부여하여 1프레임마다 학습을 한다. 본 실험에서는 제안한 NN들이 다음과 같은 4종류의 네트워크를 사용하여 학습되었다. (1) $SNR_{IN}(Input\,SNR)=20\,dB$, (2) $SNR_{IN}=15\,dB$, (3) $SNR_{IN}=10\,dB$, (4) $SNR_{IN}=5\,dB$. 학습의 실행에 필요한 각 NN의 여러 학습조건으로, 학습계수 α는 0.2, 가속도계수 β는 0.6, 초기하중은 -0.12 ~ 0.12의 난수, 입력의 실효값은 1.0으로 하여 NN을 학습시켰다. 본 실험에서는 최대 학습횟수를 오차변화가 거의 없어지는 15,000회로 하였으며, 네트워크의 구성은 20-30-20으로 하였다.

Ⅲ. 실험 결과

본 논문은 NN을 사용하여 음성신호를 강조하는 것을 목적으로 하여, 각 음성 데이터에 대한 음성강조 실험결과에 대해서 기술한다. 본 실험에서는 SNR을 20dB ~ 5dB의 환경 하에서 실시하여 본 방법의 유효성을 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)에 대하여 성능평가를 하였다. 그림 2는 백색잡음에 대하여 다양한 잡음레벨들($Input\ SNR=20\,dB\sim 0\,dB$)을 사용하여, 10개의 문장에 대한 SNR_{out}의 평균값을 나타내었다. 그림 2의 백색잡음에 대하여, 잡음이 중첩된 음성신호(Original noisy speech)와 비교하였을 때, 유성음 구간에서의 각 대역별 NN을 사용하지 않은 경우(without NN)의 SNR_{out} 최대 개선값은 약 5.5 dB, 본 방법은 약 8 dB 개선되었다. 따라서 그림에 나타낸 것과 같이 제안한 시스템은 잡음레벨이 낮았을 때보다 잡음레벨이 높았을 때에 양호한 개선결과를 보였다.

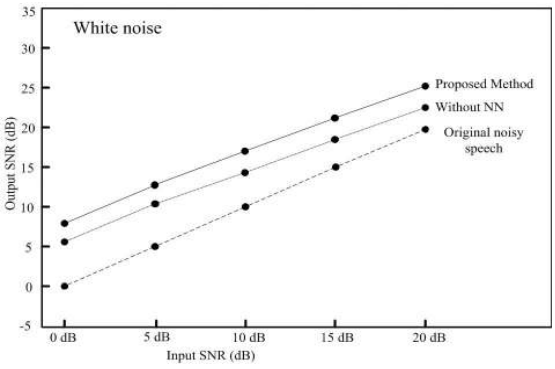

그림 2. 백색잡음 부가 시의 제안한 방식의 성능비교

Ⅳ. 결론

본 논문에서는 NN을 사용하여 잡음을 제거하는 시스템을 제안하여, 이것이 SNR에서 유효하다는 것을 백색잡음에 대해서 실험적으로 증명하였다. 따라서 제안한 시스템은 유성부 및 무성부 에 대하여 각각 저역, 중역, 고역으로 분리된 신경회로망에 의하여 잡음이 제거됨을 확인할 수 있었다. 더욱이 각 대역별 NN을 사용하지 않은 경우와 비교하여도 본 방법이 유효하다는 것을 확인할 수 있었다. 결론적으로 본 연구에서는 입력 SNR이 0 dB 정도의 조건에서도 충분히 잡음 제거 효과가 높다는 것을 확인하였다. 특히 저역부에 잡음이 집중한 유색잡음에 대해서도 저역부의 FFT 진폭성분을 복원하여 잡음을 제거할 수 있었다.

이상과 같이, 음성신호의 잡음제거를 위해서 NN에 의한 본 방식이 백색잡음에 대해서 효과적이라는 것을 실험적으로 확인하였지만, 향후의 연구과제로서는 다양한 유색잡음에 의해서 열화된 음성에 대해서도 더욱 강화하는 방법의 검토가 필요하다고 생각된다. 또한 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 잡음 하에서의 잡음억제 및 음성강조에 도움이 될 것으로 생각된다.

참고문헌

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

[2] W. G. Knecht, M. E. Schenkel, G. S. Moschytz, “Neural network filters for speech enhancement”, Transactions on Speech and Audio Processing, Vol. 3, No. 6, pp. 433-438, 1995.

[3] 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.

[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] 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.

[6] A. V. Ooyen and B. Nienhuis, "Improving the convergence of the back-propagation algorithm", Neural Networks, vol. 5, no. 3, pp. 465-471, 1992.