
신경회로망을 이용한 DFT 성분 복원에 의한 음성강조

최재승*

Noisy Speech Enhancement by Restoration of DFT Components Using Neural Network

Jae Seung Choi*

요 약

본 논문에서는 오차역전파알고리즘에 의한 신경회로망을 사용하여 이산푸리에 변환에 의한 진폭성분과 위상 성분을 복원하는 음성강조 시스템을 제안한다. 먼저, 신경회로망이 잡음이 추가된 음성신호의 이산푸리에 변환의 진폭성분과 위상성분을 사용하여 학습된 후, 제안한 시스템은 백색잡음에 의하여 열화된 잡음이 추가된 음성 신호를 강조한다. 백색잡음에 의하여 열화된 음성신호는 이산푸리에 변환에 의한 진폭성분과 위상성분을 입력으로 하는 신경회로망을 사용하여 제안된 시스템에 의하여 강조되는 것을 실험결과로 증명한다. 제안한 시스템은 스펙트럼 왜곡율의 평가법을 사용하여 백색잡음에 의하여 열화된 음성신호에 대하여 효과적인 것을 실험으로 확인한다.

ABSTRACT

This paper presents a speech enhancement system which restores the amplitude components and phase components by discrete Fourier transform (DFT), using neural network training by back-propagation algorithm. First, a neural network is trained using DFT amplitude components and phase components of noisy speech signal, then the proposed system enhances speech signals that are degraded by white noise using a neural network. Experimental results demonstrate that speech signals degraded by white noise are enhanced by the proposed system using the neural network, whose inputs are DFT amplitude components and phase components. Based on measuring spectral distortion measurement, experiments confirm that the proposed system is effective for white noise.

키워드

신경회로망, 오차역전파알고리즘, 이산푸리에 변환, 백색잡음

Key word

Neural network, back-propagation algorithm, discrete Fourier transform, white noise.

I. 서론

신경회로망이 실시하는 정보처리는 네트워크 메커니즘에 의한 병렬처리 분산형의 정보처리라고 할 수 있다. 이러한 병렬처리 분산형은 많은 단순한 정보처리 요소가 조합되어서 간단한 신호를 주고받는 네트워크 상의 메커니즘을 사용한다. 최근에 이러한 네트워크 메커니즘에 의한 정보처리가 인간의 뇌의 구조를 컴퓨터 상에서 모의하는 신경회로망에 의한 여러 연구가 활발히 시도되고 있다. 이러한 시도는 뇌와 같은 높은 병렬성을 가진 시스템에의 흥미, 자기 학습되어 가는 시스템에의 흥미 등에 지지되어 하나의 주류로서 연구의 흐름을 만들어 왔다. 이러한 연구 중에서 도출되어진 것으로는 오차역전파학습법[1, 2]에 의한 신경회로망 (Neural Network; NN)[3, 4]이 있다. 본 논문은 이러한 NN을 사용하여 음성신호에 잡음이 중첩된 신호로부터 음성신호 부분을 강조하는 것을 목적으로 하는 연구이며, 오차역전파학습법에 의한 신경회로망을 사용하여 이산푸리에변환(Discrete Fourier Transform; DFT)에 의한 진폭성분과 위상성분을 복원하는 음성강조시스템을 제안한다. 먼저, 제안한 신경회로망은 잡음이 부가된 음성신호의 DFT의 진폭성분과 위상성분을 사용하여 NN을 학습한 후에 백색잡음에 의하여 열화된 잡음이 부가된 음성신호를 강조한다. 제안한 시스템은 스펙트럼 왜곡율의 평가법을 사용하여 백색잡음에 의하여 열화된 음성신호에 대하여 효과적인 것을 실험으로 확인한다.

II. 신경회로망

1. 오차역전파 알고리즘의 학습법

오차역전파 알고리즘의 학습법[1, 2]은 Rumelhart[5]에 의해 제안된 2층오차최소화 학습을 다층 네트워크 전체의 학습에 확장한 것이다.

m 층의 네트워크를 고려하여, k 층의 제 i 유닛의 입력의 총합을 i_i^k , 출력을 o_i^k 로 하여, $k-1$ 층의 제 j 유닛으로부터 k 층의 제 j 유닛에의 결합의 가중치를 w_{ji}^{k-1k} 로 한다. 각 유닛의 입출력 관계를 다음 식과 같이 나타낸다.

$$i_i^k = \sum_j w_{ji}^{k-1k} o_j^{k-1} - \theta_i^k \quad (1)$$

$$o_i^k = f(i_i^k) \quad (2)$$

어떤 패턴의 그룹 (x, y) 이 주어질 경우의 손실함수 r 은 다음식과 같이 된다.

$$r = \sum_j (o_j^m(w, x) - y_j)^2 \quad (3)$$

네트워크의 결합의 가중치 w 의 수정량을 구하기 위해서는 다음식과 같이 r 의 w 에 대한 r 의 gradient를 계산하면 된다.

$$\Delta w_{ij}^{k-1k} = -\varepsilon \frac{\partial r}{\partial w_{ij}^{k-1k}} \quad (4)$$

$$\frac{\partial r}{\partial w_{ij}^{k-1k}} = \frac{\partial r}{\partial i_j^k} \frac{\partial i_j^k}{\partial w_{ij}^{k-1k}} = \frac{\partial r}{\partial i_j^k} o_i^{k-1} \quad (5)$$

f' 는 f 의 도함수가 성립되기 때문에, $\frac{\partial r}{\partial i_j^k} = d_j^k$ 로 두면, 결합 수정량 Δw_{ij}^{k-1k} 는 다음 식을 만족한다.

$$\Delta w_{ij}^{k-1k} = -\varepsilon d_j^k o_i^{k-1} \quad (6)$$

$$d_j^m = (o_j^m - y_j) f'(i_j^m) \quad (7)$$

$$d_j^k = \left(\sum_i w_{ji}^{k+1k} d_i^{k+1} \right) f'(i_j^k) \quad (8)$$

ε 은 1회의 수정의 크기를 결정하는 파라미터이며 작은 정의 정수이다. 여기에서 f 가 모든 유닛에서 공통으로 식 (9)의 로지스틱 함수를 만족한다.

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (9)$$

도함수 $f'(x)$ 및 $f'(i_j^k)$ 는 식과 같이 표현한다.

$$f'(x) = f(x)(1-f(x)) \quad (10)$$

$$f'(i_j^k) = o_j^k(1-o_j^k) \quad (11)$$

실제로는 진동을 감소시키고 학습의 수속을 빨리하기 위해서는 식 (12)와 같이 되는 수정을 추가하는 경우도 있다.

$$\Delta w_{ij}^{k-1k}(t+1) = -\epsilon d_j^k o_i^{k-1} + \alpha \Delta w_{ij}^{k-1k}(t) \quad (12)$$

여기에서 t 는 수정의 횟수, ϵ 는 학습계수, α 는 가속도계수이다.

2. 실험에 사용한 네트워크의 구조

실험에 사용한 신경회로망은 그림 1에 나타내는 NN이며, 입력층이 17유닛, 중간층1이 17유닛, 중간층2가 17유닛, 출력층이 17유닛을 가지는 구조를 가진다.

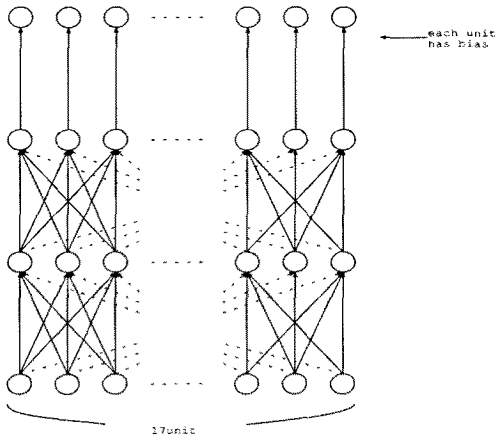


그림 1. 제안한 NN 시스템의 구조
Fig. 1. Construction of proposed NN system.

본 실험에서는 32 포인트의 DFT를 실시하므로 실제로 출력되는 계수는 32개이지만, 음성신호 샘플값이 실수 데이터이기 때문에 계수를 $X(n)$, $0 \leq n \leq 31$ 과 같이 하면, 이 계수 사이에는 다음 식의 관계가 성립된다.

$$X(i) = X(32-i)^*, \quad 0 \leq i \leq 15 \quad (13)$$

따라서 실제로 입력에 사용되는 DFT 계수는 17개면 충분하다. 또한, 본 네트워크의 학습방법은 제II장에서 기술한바와 같은 오차역전파 학습법(Back-Propagation;

BP)을 사용한다. 그리고 이후의 실험에서 NN의 학습횟수는 10,000회로 하여 실험을 실시하였다.

III. NN을 이용한 DFT 계수 복원에 의한 음성강조

본 장에서는 NN의 학습을 이용하여 잡음이 중첩된 DFT 계수의 값을 잡음이 중첩되기 전의 값으로 복원하는 실험에 대하여 기술한다.

1. 실험 데이터 및 실험 원리

본 실험에서 사용한 음성데이터는 일본인 남성화자에 의한 단어 "aioi(M1)", "hachioji(M2)", "bizen(M3)"를 사용하였으며, 샘플링 주파수는 8 kHz이며 양자화 비트수는 12비트이다. 그리고 백색잡음은 컴퓨터에 의해서 작성한 가우스 잡음을 사용하였으며 샘플링 주파수는 8 kHz이다.

본 실험에 있어서의 평가방법으로서는 스펙트럼 왜곡율(Spectral Distortion; SD)을 사용한다. SD는 입출력 신호의 대수 스펙트럼의 차를 구한 것으로 식 (14)와 같이 정의한다.

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{1}{W} \int_0^w S_x(w) - S_y(w)^2 dw} \quad (14)$$

단, $S_x(w)$, $S_y(w)$ 는 입출력 신호의 대수 스펙트럼 (dB)이며, N 은 측정구간의 프레임 수, W 는 신호의 대역폭이다.

본 실험에서는 1프레임을 32 샘플로 하는 음성신호 표본값을 이산푸리에 변환하여 DFT 계수를 구한다. NN의 입력신호로는 백색잡음을 부가한 음성신호로부터 구한 DFT 계수를, 교사신호로서는 잡음을 부가하지 않은 음성신호로부터 구한 DFT 계수를 각각 부여하여, BP법에 의하여 NN의 학습을 실시한다. 단, DFT 계수는 실수부와 허수부로 분리하여 각각 별도로 NN을 구축하여 학습을 실시한다. 이후에, 잡음을 부가한 음성신호로부터 구한 DFT 계수를 학습 종료 후의 NN에 부가하여 출력력을 구한다. 이 출력으로서 구해진 DFT 계수를 역이산 푸리에 변환(Inverse Discrete Fourier Transform; IDFT)하

여 음성신호의 샘플값을 구한다. 이 신호와 잡음을 부가하기 전의 음성신호 샘플값을 비교하는 것에 의하여 효과를 조사한다.

2. NN의 정규화 방법 및 학습방법

본 실험에서는 NN에의 입력인 DFT 값을 0부터 1까지의 범위로 정규화하며, 본 실험에서 사용한 데이터의 정규화 방법은 다음과 같다. DFT 계수의 실수부 및 허수부의 각 값은 30,000을 부가하여 60,000으로 나눈다. 이 이유는 각 값이 -30,000부터 30,000의 범위에 있다는 가정을 기본으로 하기 때문이다. 실제로 값의 상한은 20,000을 초과하는 것은 없기 때문에 이 방법에 의하여 데이터를 0부터 1의 범위에 제한이 가능하다. 이와 같이 하여 정규화된 각 프레임의 DFT 계수는 입력층의 각 유닛에 부여되기 때문에 다음과 같은 방법으로 학습이 실시된다. (1) 각각의 유닛의 가중치 및 문턱치를 임의로 0부터 1까지의 범위 내에서 초기화하여 학습계수(II.2절의 식 12)를 지정한다. (2) 잡음이 중첩된 음성신호의 제1프레임으로부터 구한 DFT 계수를 실수부 및 허수부로 분리하여, 각각 별도의 신경회로망의 입력층의 각 유닛에 부여한다. (3) 잡음이 부가되지 않은 음성신호의 제1프레임으로부터 구한 DFT 계수를 실수부 및 허수부로 분리하여, 각각의 NN에 교사신호로써 부여한다. (4) II.2절에서 기술한 BP 알고리즘에 의하여 학습을 실시한다. (5) 동일한 방법으로 다음 프레임의 DFT 계수를 입력신호 및 교사신호로써 부여한다. (6) 이하, 동일한 방법의 조작을 반복하여 학습을 실시하며, 모든 프레임에 대하여 학습이 종료된 시점에서 (2)로 되돌아간다. 단, 되돌아 가는 횟수는 학습횟수로서 지정된 횟수로 한다. 이와 같은 방법으로 학습 후의 네트워크에 잡음이 중첩된 음성신호로부터 구한 DFT 계수를 입력신호로써 부여하면 출력으로써 DFT 계수가 구해지게 된다.

IV. 실험결과

본 장에서는 NN을 사용하여 잡음이 부가된 음성신호를 강조하는 것을 목적으로 한 실험결과에 대해서 기술한다. 학습에 사용한 파라미터는 ϵ (학습계수): 0.3, α (가

속도계수): 0.3, 수렴판정오차: 1.0×10^{-4} , 최대학습횟수: 10,000회이다. NN 학습의 일반성 및 잡음강도를 변경하여도 효과가 구해지는가를 조사하기 위하여, 표 1은 학습 시에 입력으로써 사용한 데이터와는 잡음강도가 다른 음성신호를 학습 후의 네트워크에 입력함으로써 출력을 구한 경우의 효과를 나타낸다. 표 1의 결과로부터 학습 시 입력으로써 사용한 음성신호에 부가되었던 잡음과는 다른 강도의 잡음에 대해서도 SD 경감의 효과가 구해지는 것을 알 수 있다.

그림 2, 3, 4는 백색잡음이 DFT 계수의 값에 미치는 영향을 조사하기 위하여 NN의 학습에 의한 DFT 계수의 변화에 대한 학습효과를 예들 각각 나타낸다. 그림 2는 잡음이 없는 "bizen"의 제100프레임 DFT 계수의 절대값을, 그림 3은 잡음이 부가된 SD=13.74의 "bizen"의 제100프레임의 DFT 계수의 절대값을, 그림 4는 학습 후의 "bizen"의 제100프레임의 DFT 계수의 절대값을 각각 나타낸다. 여기에서 가로축은 샘플수, 세로축은 DFT 계수의 진폭의 절대값을 각각 나타내며, 해당 100 프레임은 임의적으로 선택하였다. 그림들을 보면, 잡음이 부가된 음성의 DFT 계수가 학습에 의하여 잡음이 부가되기 전의 DFT 계수에 근접하는 것을 알 수 있다.

표 1. 학습 시의 입력신호에 강도가 다른 잡음을 부가한 경우의 실험결과

Table 1. Experiment results when training and estimating noisy intensity were different.

Training		After training			
Input speech	Input SD(dB)	Input speech	SD(dB)		
			Input	Output	Impr.
M1	10.47	M1	8.22	7.30	0.92
		M1	12.56	9.77	2.79
		M1	14.68	11.55	3.13
M2	11.94	M2	10.55	8.78	1.77
		M2	14.50	10.10	4.40
		M2	16.18	11.27	4.91
M3	13.74	M3	10.70	7.50	3.20
		M3	16.52	10.39	6.13
		M3	18.57	12.23	6.34

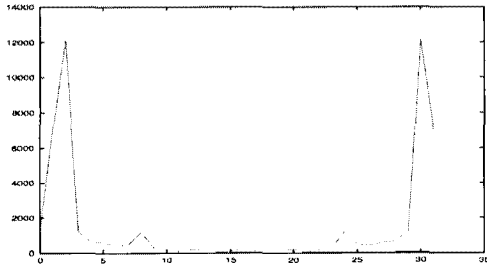


그림 2. 음성신호의 DFT 계수의 절대값
Fig. 2. Absolute values of DFT coefficient of speech signal.

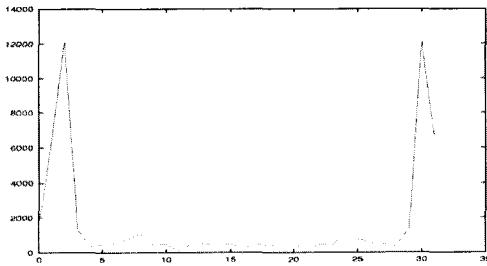


그림 3. 잡음이 추가된 음성의 DFT계수 절대값
Fig. 3. Absolute values of DFT coefficient of noisy speech signal.

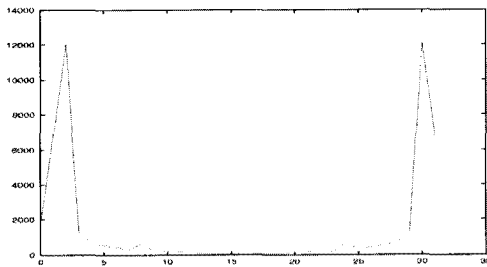


그림 4. 학습후의 음성신호의 DFT계수 절대값
Fig. 4. Absolute values of DFT coefficient of noisy speech signal after training.

표 2는 음성신호를 변경하여도 본 논문에 의한 효과가 구해지는가를 조사하기 위하여, 학습 시에 입력으로 사용한 음성과는 다른 음성에 백색잡음을 추가한 것을 학습 후의 NN에 입력하여 출력을 구하는 실험을 하였다. 표 2의 결과로부터 알 수 있듯이 다른 음성에 대해서도 NN에 의하여 DFT 계수를 잡음이 추가되기 전의 상태로 근접시킴으로써 SD 경감의 효과가 보이는 것을 알 수

있다.

그림 5는 음성신호 "aioi"의 제3프레임에 대한 NN의 학습신호의 DFT 진폭성분을, 그림 6은 잡음이 추가된 SD=10.47의 음성신호 "aioi"의 제3프레임에 대한 학습하지 않은 시점에서의 NN의 입력신호의 DFT 진폭성분을, 그림 7은 잡음이 추가된 SD=10.47의 음성신호 "aioi"의 제3프레임에 대한 학습한 시점에서의 NN 학습 후의 출력신호의 DFT 진폭성분을 각각 나타낸다. 여기에서 가로축은 샘플수, 세로축은 DFT 계수의 진폭을 각각 나타내며, 해당 3 프레임은 임의적으로 선택하였다. 그림들을 비교함으로써 "aioi"를 학습함으로써 DFT 진폭성분이 충분히 잡음이 없는 상태에 접근한 것을 알 수 있다.

이상의 결과로부터 입력 SD가 최대 18(dB) 정도까지의 잡음에 대해서도 본 방식에 의한 잡음제거 효과가 높은 것을 확인할 수 있었다.

표 2. 학습 시의 입력신호와 다른 음성신호를 사용한 경우의 실험결과

Table 2. Experiment results when training and estimating speech data were different.

Training		After training			
Input speech	Input SD(dB)	Input speech	SD(dB)		
			Input	Output	Impr.
M1	10.47	M2	11.94	10.36	1.58
		M2	14.50	12.24	2.26
		M2	16.18	13.74	2.44
		M3	13.71	10.94	2.77
		M3	16.52	13.31	3.21
		M3	18.57	15.42	3.15
M2	11.94	M1	10.47	8.22	2.25
		M1	12.56	9.16	3.40
		M1	14.68	10.45	4.23
		M3	13.71	8.89	4.82
		M3	16.52	10.81	5.71
		M3	18.57	12.54	6.03
M3	13.74	M1	10.47	8.68	1.79
		M1	12.56	9.42	3.14
		M1	14.68	10.22	4.46
		M2	11.94	9.38	2.56
		M2	14.50	10.41	4.09
		M2	16.18	11.43	4.75

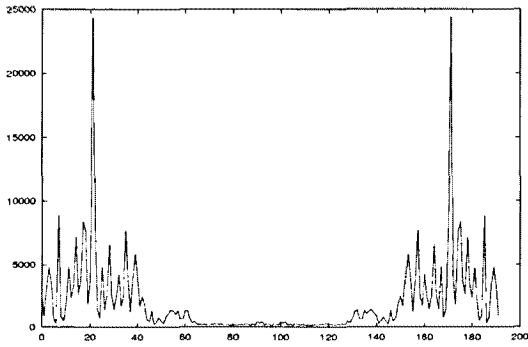


그림 5. NN의 학습신호의 DFT 진폭성분
Fig. 5. DFT amplitude components for training signal.

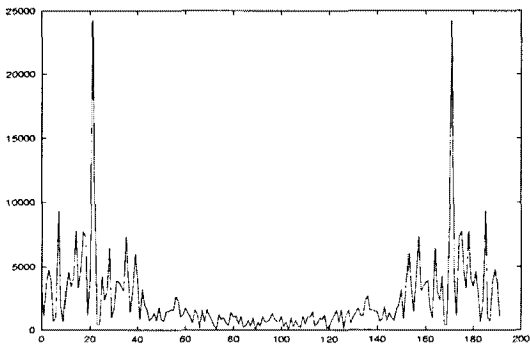


그림 6. NN의 입력신호의 DFT 진폭성분
Fig. 6. DFT amplitude components for input signal.

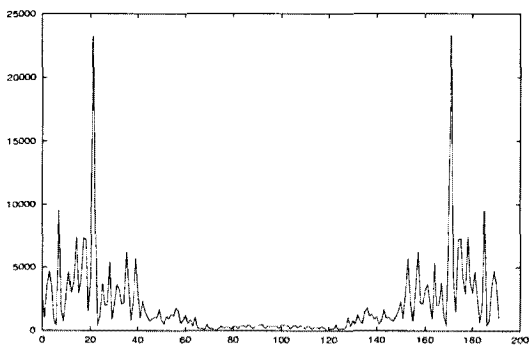


그림 7. 학습 후의 출력신호의 DFT 진폭성분
Fig. 7. DFT amplitude components for input signal after trained.

V. 결론

본 논문에서는 음성강조의 방법으로써 BP법에 의한 NN을 사용하여 DFT에 의한 진폭 및 위상성분을 복원하는 음성강조시스템을 제안하였다. 본 시스템은 NN의 학습에 의하여 잡음이 부가된 음성신호의 진폭 및 위상성분을 잡음이 부가되기 전의 음성신호로 복원 가능한 것을 SD평가법을 사용하여 효과적인 것을 실험으로 확인하였다.

향후의 연구과제로는 (1) NN은 실수부 및 허부수를 별도로 학습하고 있기 때문에 학습에 걸리는 시간을 단축할 필요가 있다. (2) 실험에 사용하는 음성데이터를 다양하게 할 필요가 있다.

이상으로 NN에 의한 DFT 계수의 복원에 의한 본 논문의 수법이 스펙트럼 왜곡율의 값에 의한 개선이 음성신호처리의 분야에 많은 도움이 될 것으로 생각한다.

참고문헌

- [1] Ooyen A. V. and Nienhuis B. "Improving the convergence of the back-propagation algorithm," *Neural Networks* 5, 3, pp. 465-471, 1992.
- [2] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by back-propagation errors," *Nature* 323, pp. 533-536, 1986.
- [3] K. Daqrouq, I. N. Abu-Isbeih, M. Alfauri, "Speech signal enhancement using neural network and wavelet transform", 2009 6th International Multi-Conference on Systems, Signals and Devices, pp. 1-6, 2009.
- [4] W. G. Knecht, M. E. Schenkel, G. S. Moschytz, "Neural network filters for speech enhancement", *IEEE Trans. Speech and Audio Processing*, Vol. 3, No. 6, pp. 433-438, 1995.
- [5] D. Rumelhart, "Parallel Distributed Processing, vol. 1 and 2, MIT Press, Cambridge, MA, 1986.

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~2007년 경북대 디지털기술연구소 책임연구원

2007년~현재 신라대학교 전자공학과 교수

※관심분야: 디지털통신, 음성신호처리, 신경회로망,
적응필터와 잡음제거, 디지털 TV 등