

논문 2005-42SP-3-15

시간지연 신경회로망을 이용한 잡음제거 시스템

(Noise reduction system using time-delay neural network)

최 재 승*

(Jae-Seung Choi)

요 약

음성신호를 대상으로 하는 연구 분야에서 신경회로망은 주로 음성인식 등의 카테고리 분류의 목적으로 사용되며 신호처리의 응용에도 유망하다. 따라서 본 논문에서는 신경회로망에 시간구조를 취한 시간지연 신경회로망을 이용하여 잡음이 중첩된 음성신호의 공간으로부터 잡음이 없는 음성신호의 공간으로 사상을 실행함으로써 잡음을 제거하는 것을 목적으로 한다. 본 논문은 푸리에 변환의 진폭성분을 복원하는 잡음제거의 알고리즘을 사용하여 백색잡음 및 유색잡음에 대해서 본 수법의 유효성을 확인한다.

Abstract

On the research field for speech signal, neural network mainly uses for the category classification in speech recognition and applies to signal processing. Accordingly, this paper proposes a noise reduction system using a time-delay neural network, which implements the mapping from the space of speech signal degraded by noise to the space of clean speech signal. It is confirmed that this method is effective for speech degraded not only by white noise but also by colored noise using the noise reduction system, which restores the amplitude component of fast Fourier transform.

Keywords : Signal processing, Speech recognition, Noise reduction, Time-delay neural network.

I. 서 론

음성인식 등의 음성정보처리의 실용화를 위한 중요한 문제점으로 실제의 잡음 환경 하에 대한 대응이 필요하며 이에 대한 여러 연구가 진행되고 있다^{[1][2]}.

음성강조 및 잡음제거를 위하여 신경회로망 (neural network: NN)^{[3][4]}, 스펙트럼 차감법 (spectral subtraction)^{[5][6]}, 위너필터 (Wiener filter)^[7], microphone array^{[3][8]}, 적응 필터법^{[9][10]} 등의 방식이 발표되었다. 이러한 논문 중에 NN이 음성신호처리 분야에 사용되는 경우의 주요한 응용분야는 음운인식 등의 카테고리적인 분류이며, 음성강조 및 잡음제거의 목적으로 사용되는 경우에는 잡음 중의 음성구간의 추출^{[11][12]} 등의 전 처리적인

역할을 담당하는 경우가 많다. 그러나 다층 퍼셉트론이 원리적으로 임의의 연속 사상(mapping)을 실현 할 수 있다는 것과 역전파(back propagation : BP) 알고리즘이라는 간단하고 강력한 알고리즘이 있다는 것을 생각하면, NN이 음성신호처리 분야에의 응용도 유망하다고 고려된다. 따라서 본 논문은 NN을 사용하여 잡음이 중첩된 음성신호의 공간으로부터 잡음이 없는 음성신호의 공간으로 사상을 실행함으로써 잡음을 제거하는 것을 목적으로 한다.

음성신호의 관점으로부터 본 경우의 음성신호는 시간변화가 중요한 정보 중의 하나이며, NN을 이용하여 음성신호를 처리하는 경우도 시간구조를 NN에 구성하는 점도 중요하다. 따라서 본 논문에서는 NN에 시간요소를 도입한 시간이 지연된 신경회로망(time-delay neural network: TDNN)^[13]을 사용한다. 또한 음성신호를 고속 푸리에 변환(fast Fourier transform : FFT)한

* 정회원, 경북대학교 디지털기술연구소
(Digital Technology Research Center, Kyungpook National University)
접수일자: 2005년1월4일, 수정완료일: 2005년3월17일

경우 위상성분보다 진폭성분이 음성 정보를 많이 포함하고 있다. 따라서 본 논문은 스펙트럼 회복의 수단으로써 TDNN을 이용하여 FFT 진폭성분을 복원하는 알고리즘을 제안한다. 본 알고리즘을 사용하여 음성신호에 대한 잡음제거의 실험을 하였으며 본 실험결과와 위너필터를 사용한 경우의 실험결과를 비교하여 본 알고리즘의 유효성을 확인한다.

II. 일반적인 시간지연 신경회로망의 구조

음성신호 등을 입력으로 하여 NN을 학습하는 경우에 신호의 시간변화가 중요한 요소로 작용하는 경우가 많다. 이러한 사항을 고려한 NN으로 TDNN이 고안되어 있으며 본 논문에서는 TDNN을 사용한다. 일반적인 TDNN의 구조는 그림 1과 같다. 입력 층의 가

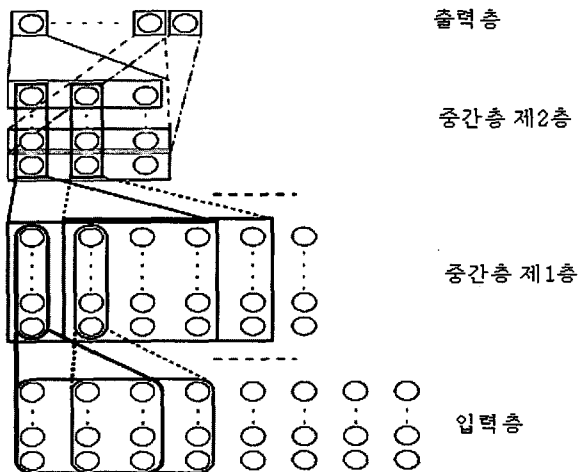


그림 1. 일반적인 TDNN의 구조
Fig. 1. Construction of typical TDNN.

표 1. 사용한 화자와 음성 데이터와의 관계
Table 1. Relation between speakers and speech data.

화자	음성
남성화자 A	"aioi", "bizen", "yatiyo", "kisarazu", "hikone", "sendai", "hachioji", "shimonoseki"
남성화자 B	M1: "tsukubawa donna tokorodesuka", M2: "arigatou gozaimashita" M3: "kondono onsei kenkyuukai", M4: "kanazawani ryokoushitainode"
여성화자 C	F1: "toukyou tawano mae", F2: "kaijouwa dochiradesuka"

로축 방향은 시간을 나타내며, 여기에서 시간 축 방향의 최소단위를 1프레임이라고 한다. 그림은 입력층의 3프레임 분의 유닛(unit)은 중간층 제1층의 1프레임분의 유닛에 결합된다. 이러한 결합에 의해서 신호의 단기간 시간 변화 패턴이 중간층 제1층에 반영된다. 또한 중간층 제1층의 4프레임분의 유닛은 중간층 제2층의 1프레임분의 유닛과 결합된다. 이 결과, 장기적인 시간변화 패턴이 중간층 제2층의 유닛에 반영된다. 더욱이 중간층 제2층의 각 가로 1열의 유닛은 출력층의 각 유닛에 대응하여 접속되어 있다. 이러한 구조를 가짐으로써 신호의 시간변화 패턴을 반영할 수 있는 네트워크가 구축 가능하다.

III. 실험조건

1. 음성신호

본 실험에 사용한 음성데이터는 일본인 남성 및 여성 화자가 발성한 단어 또는 문장을 사용하였으며, 모두 샘플링 주파수는 8kHz이다. 표 1은 음성 데이터와 화자와의 관계를 나타내며, 화자 B, C의 음성 데이터는 일본정보처리 개발협회에서 배부한 연구용 연속음성 데이터베이스 중의 문장이다. 표에서 M은 남성화자를 나타내며, F는 여성화자를 나타낸다. 음성 "aioi", "bizen", "yatiyo", "kisarazu"에는 백색잡음을, 음성 "hikone", "sendai", "M1", "M2"에는 유색잡음을 추가하여 TDNN을 학습시켰으며, 나머지 음성은 학습결과의 평가에 사용하였다.

본 실험에서는 음성신호를 $s(n), n=1, 2, \dots, N, F$ 를 어떤 정수라고 한다면, 식 (1)과 같이 되도록 전력에 대해서 정규화를 실시하였다.

$$\sum_{n=1}^N \frac{s^2(n)}{N} = F \tag{1}$$

2. 잡음신호

본 실험에 사용한 배경잡음으로는 (1) 컴퓨터에 의해서 생성한 가우스(gauss) 백색잡음, (2) 음성과 비슷한 스펙트럼을 가진 유색잡음의 2종류를 사용하였다. 유색잡음은 그림 2의 진폭특성을 가진 저역통과필터에 백색잡음을 통과시킴으로써 생성하였다. 그림에서 알 수 있듯이 유색잡음의 스펙트럼은 음성의 스펙트럼과 거의 유사한 주파수 분포를 하고 있기 때문에 이러한 음성과 유사한 잡음을 제거하는 것은 쉽지 않다. 결과로서 구

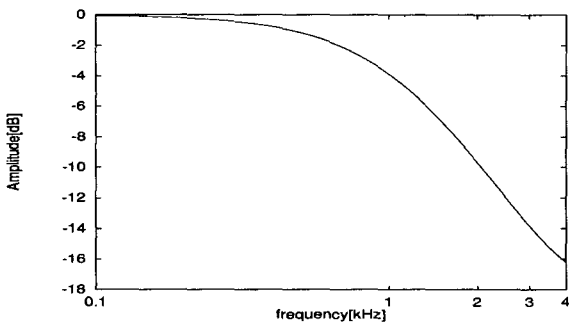


그림 2. 저역통과 필터의 진폭 특성
Fig. 2. Amplitude characteristic of low-pass filter.

해지는 유색잡음의 스펙트럼은 음성의 스펙트럼에 상당히 비슷하기 때문에 NN으로 잡음과 음성을 식별하는 것은 곤란하다고 생각되어지지만, 본 실험에 의해서 충분히 식별 가능하다는 것을 나타낸다. 이러한 배경잡음으로 사용한 데이터는 모두 샘플링 주파수 8kHz이다. 본 실험에서의 음성 및 잡음의 주파수 대역은 0~3kHz이다.

IV. 제안한 TDNN

1. TDNN의 구조

TDNN의 입력신호에는 잡음이 중첩된 음성신호로부터 구해진 FFT 진폭성분이 부여되며 학습신호에는 잡음을 부가하지 않은 음성신호로부터 구해진 FFT 진폭성분을 부여하여 1프레임마다 학습을 한다. 단, TDNN에는 부가적인 정보로서 학습대상 프레임의 이전 2 프레임과 다음 1프레임이 부여된다. 따라서 입력되는 총 프레임 수는 4 프레임이다. 본 논문에서는 크게 떨어져 있는 주파수 대역의 스펙트럼 간에 상관이 적은 정보를 TDNN에 제공하는 것이 학습의 효과가 높다고 판단하여 저역부용 및 고역부용에 각각의 TDNN을 구축하며, 또한 어느 정도 비슷한 패턴마다 TDNN을 구축하는 경우가 학습하기 용이하다고 판단하여 유성부용 및 무성부용에 각각의 TDNN을 구축하는 그림 3의 개량된 TDNN 시스템을 제안한다. 먼저 입력음성신호는 유성/무성으로 판별되어 FFT된 후에 저역, 고역의 FFT 성분으로 분할되어 각각의 TDNN에 부가된다.

유성/무성의 판별은 유성부/무성부의 대략적인 판별을 실행하는 것이 목적이며 본 실험에서는 상세한 정밀도는 필요하지 않다고 판단하여 다음과 같은 방법으로 유성음/무성음으로 판별한다. (1) 음성신호를 256 샘플

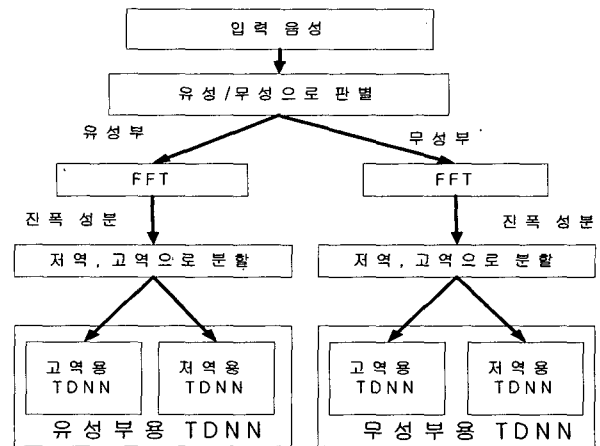


그림 3. 제안한 TDNN 시스템
Fig. 3. Proposed TDNN system.

의 프레임으로 분할한다. (2) 분석대상의 프레임에 대해서 해밍창(hamming window)를 곱한다. (3) 이후, FFT하여 저역부(1kHz 이하)에 일정한 문턱값 이상의 전력 스펙트럼의 값이 존재하면 이 프레임은 유성부로, 존재하지 않으면 무성부로 판단한다. FFT는 유성/무성으로 판별된 후의 음성신호에 대해서 1프레임 128샘플로 FFT를 실행한다. FFT를 한 후 구해진 진폭성분은 63번째의 성분을 중심으로 대칭한 값이므로 용장(redundancy)부를 제외한 0~63샘플을 저역 및 고역으로 분할하여, 각각을 저역용 TDNN, 고역용 TDNN에의 입력으로 한다. 각 TDNN의 입력에는 저역용 TDNN에는 0~31샘플(0~1.9kHz), 고역용 TDNN에는 32~63샘플(2kHz~3.9kHz)을 입력신호로 한다.

2. 학습 방법

본 논문에서 제안한 저역용 TDNN 및 고역용 TDNN은 그림 4와 같은 구조를 가지며 BP 알고리즘에 의해서 학습이 실행된다. 저역용 TDNN의 입력신호는 저역용 TDNN에의 입력신호로서 부여된 FFT 진폭성분 0~31샘플(이 프레임을 학습대상 프레임이라 부르며 제n프레임이라 한다)과 부가정보적인 프레임으로서 제n프레임의 이전 2프레임(제n-1, 제n-2프레임) 및 다음 1프레임(제n+1프레임)을 FFT함으로써 구해진 FFT 진폭성분 0~31샘플이다. 학습신호는 잡음을 부가하지 않은 음성신호 중, 학습대상 프레임에 해당하는 프레임을 FFT함으로써 구해진 FFT 진폭성분 0~31샘플이다. 고역용 TDNN의 입력신호는 고역용 TDNN에의 입력신호로서 부여된 FFT 진폭성분 32~63샘플과 부가정보

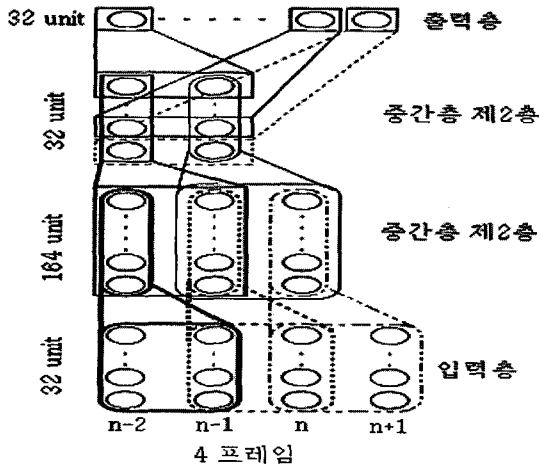


그림 4. 제안한 저역 및 고역용 TDNN의 구조
 Fig. 4. Construction of proposed TDNN for low and high frequency.

표 2. TDNN 학습시의 여러 조건
 Table 2. Various conditions for training of TDNN.

학습계수	$\alpha=0.1$
가속도계수	$\beta=0.6$
학습제한 횟수	10,000회
초기하중	-0.06 ~ 0.06의 난수
입력의 실효값	1.0

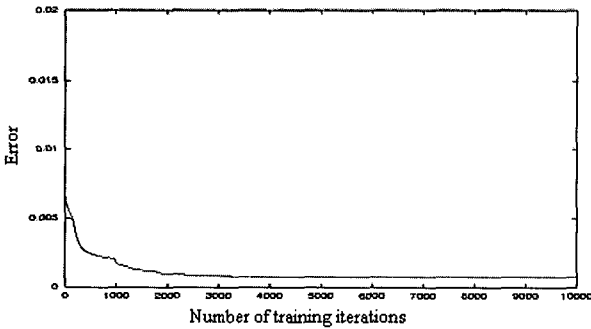


그림 5. 학습 시에 대한 오차 곡선
 Fig. 5. Error curves of training.

적인 프레임으로서 구해진 FFT 진폭성분 32~63샘플이다. 학습신호는 잡음을 부가하지 않은 음성신호 중, 학습대상 프레임에 해당하는 프레임을 FFT하는 것에 의해서 구해진 FFT 진폭성분 32~63샘플이다.

3. TDNN의 학습조건

학습의 실행에 필요한 각 TDNN의 여러 학습조건을 표 2에 나타낸다. 본 실험에서는 최대 학습 횟수를 오차변화가 거의 없어지는 10,000회로 하였다.

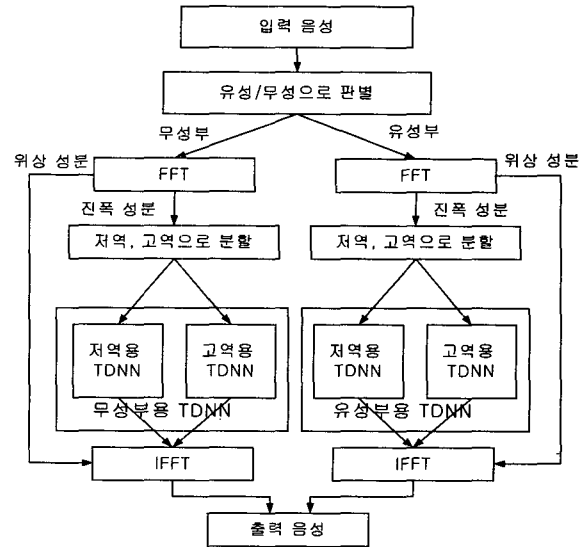


그림 6. 학습 후의 TDNN의 이용
 Fig. 6. Use of TDNN after trained.

그림 5는 화자 A가 발생한 신호대잡음비(Signal-to-noise ratio: SNR)가 약 -10dB의 음성신호 "aioi"를 각 TDNN에 부여해서 학습을 하였을 경우의 유성부에 대한 저역용 TDNN에서의 오차곡선을 나타낸다. 여기에서 오차곡선은 하나의 네트워크에 대해서 10회의 시행에 대한 평균값이다.

4. 학습 후의 TDNN의 이용방법

그림 6은 학습 후의 TDNN의 이용 방법을 나타내며, 각 TDNN의 입력에 도달할 때까지는 그림 3의 학습 시의 TDNN 시스템과 동일하다. 먼저 입력음성신호는 유성/무성으로 판별되어 FFT된 후에 저역, 고역의 FFT 성분으로 분할되며, 각각의 TDNN에 부가되어 출력된다. 각각의 저역용 TDNN 및 고역용 TDNN으로부터의 출력을 합성하여 구해진 FFT 진폭성분과 FFT부에서 구해진 위상성분을 역 고속 푸리에 변환 (inverse fast Fourier transform : IFFT)하여 출력 음성신호를 구한다.

V. 실험결과

본 논문은 TDNN을 사용하여 음성신호에 중첩된 잡음을 제거하는 것을 목적으로 하여, 음성 데이터에 대한 잡음제거의 실험결과에 대해서 기술한다. 본 실험에서는 모두 SNR이 약 -10dB의 환경 하에서 실시하여 본 방법의 유효성을 스펙트럼 왜곡척도(spectral distortion: SD)를 사용하여 확인한다.

1. 실험결과의 평가법

본 논문에서 제안한 잡음제거의 방법과 비교하기 위하여 최적필터의 일종인 위너 필터를 사용한다. 또한 음성품질을 평가하기 위하여 본 논문에서는 시간영역의 평가척도인 SNR, 주파수영역의 평가척도인 SD를 사용한다. SNR의 척도를 식 (2)와 같이 정의한다.

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^N X(i)^2}{\sum_{i=1}^N X(i) - Y(i)^2} \quad (2)$$

여기에서, $X(i)$, $Y(i)$ 는 각각 입력신호 및 출력신호의 표본값이며 N 은 측정구간의 표본수를 나타낸다.

SD의 척도는 입출력신호의 대수 스펙트럼의 차를 구한 것이며 연속함수 $x(t)$, $y(t)$ 에 대해서 식 (3)과 같이 정의한다.

$$SD = \sqrt{\frac{1}{N_F} \sum_{i=1}^{N_F} \frac{1}{W} \int_0^W S_x^{(i)}(f) - S_y^{(i)}(f)^2 df} \quad (3)$$

여기에서, N_F 는 측정구간의 프레임 수, W 는 신호의 대역폭, $S_x^{(i)}(f)$ 및 $S_y^{(i)}(f)$ 는 입력신호 $x(t)$ 와 출력신호 $y(t)$ 의 대수 스펙트럼(dB)이며, 식 (4)와 같이 정의한다.

$$\begin{aligned} S_x^{(i)}(f) &= 10 \log_{10} |X(f)|^2 \\ S_y^{(i)}(f) &= 10 \log_{10} |Y(f)|^2 \end{aligned} \quad (4)$$

여기에서, $X(f)$, $Y(f)$ 는 각 주파수 f 에서의 입력신호와 출력신호의 스펙트럼이다. 이와 같이 각 프레임의 SD를 구하여 이 평균값을 측정구간 전체의 SD로 한다.

2. 백색잡음에 대한 효과

학습에 사용한 음성에 대한 효과를 조사하기 위하여 학습 시의 입력신호와 동일한 음성신호에 백색잡음을 부가하여 학습 시의 TDNN에 입력함으로써 표 3의 입력신호와 출력신호에 대한 SD값을 구하였으며, 위너 필터를 사용한 수법과도 비교를 하였다. 위너 필터는 TDNN의 학습에 사용한 음성 전체로부터 구한 상관계수를 사용하였으며 필터의 계수는 선형예측으로 사용된 차수를 고려하여 10차로 하였다. 표 3의 M1에 백색잡음이 부가된 경우를 예로 들면, 위너필터에 의한 SD값

표 3. 학습한 음성에 대한 입력 및 출력신호의 SD

Table 3. SD values of input and output signal for trained speech.

입력 음성	입력신호		출력신호 SD(dB)	
	SNR(dB)	SD(dB)	위너필터	본 방식
"aioi"	-10.07	17.52	11.70	8.02
"bizen"	-10.05	21.49	15.54	11.35
"yatiyo"	-10.08	16.92	11.51	8.20
"kisarazu"	-10.10	18.63	12.83	9.69
"hikone"	-10.12	18.95	13.21	10.06
"sendai"	-10.02	19.16	13.33	9.51
M1	-10.04	21.59	14.67	10.03
M2	-10.09	20.42	13.72	9.57

표 4. 학습하지 않은 음성에 대한 입력 및 출력신호의 SD

Table 4. SD values of input and output signal for non-trained speech.

입력 음성	입력신호		출력신호 SD(dB)	
	SNR(dB)	SD(dB)	위너필터	본 방식
"hachioji"	-10.06	18.61	13.73	9.87
"shimonoseki"	-10.09	19.15	14.62	10.46
M3	-10.08	19.56	14.42	9.69
M4	-10.10	21.17	15.71	10.89
F1	-10.05	21.12	15.73	11.20
F2	-10.07	20.86	15.19	10.45

은 14.67(dB)이며 본 방식에 의한 SD값은 10.03(dB)이다. 이것은 백색잡음을 포함한 원음에 대한 SD 값인 21.59(dB)보다 위너필터에 대해서는 6.92(dB), 본 방식에 대해서는 11.56(dB)이 각각 개선되었다. SD의 절대치에 의한 평가로부터 SNR의 평균 방법과 다르기 때문에 단순히 SNR과 비교하기는 어렵지만, 전화의 PCM 방식의 평가 데이터에 의하면 SNR에서의 20(dB)의 개선량은 SD에서는 약 3.5(dB)에 대응하므로(참고 문헌^[14]의 Fig. 3), SD에서의 11.56(dB)의 개선량은 상당히 효과적이라고 말할 수 있다. 실제로, 출력된 음성을 들어본 결과, SD가 적을수록 양호한 재생 음성임을 확인할 수 있었다.

TDNN의 일반성을 조사하기 위하여 TDNN에 학습하지 않은 음성신호를 부여하여 표 4의 입력신호와 출력신호에 대한 SD를 구하였으며 F2에 대해서 SD가 최대 10.41(dB) 개선되었다. 이 표들로부터 학습한 음성신

표 5. 학습하지 않은 음성에 대한 입력 및 출력신호의 SD

Table 5. SD values of input and output signal for non-trained speech.

입력음성	입력신호		출력신호 SD(dB)	
	SNR(dB)	SD(dB)	위너필터	본 방식
"hachioji"	-10.08	16.73	12.65	8.31
"shimonoseki"	-10.12	17.70	13.77	9.29
M3	-10.10	17.38	13.08	8.28
M4	-10.12	18.57	14.08	9.07
F1	-10.08	18.52	14.14	9.12
F2	-10.09	18.40	13.89	8.70

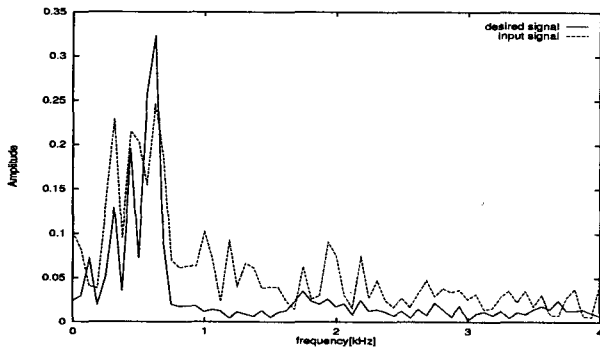


그림 7. 학습신호와 입력신호의 FFT 진폭성분의 비교
Fig. 7. Comparison of FFT amplitude component for training signal and input signal.

호뿐만 아니라 학습하지 않은 음성신호에 대해서도 잡음제거의 효과가 구해졌다는 것을 알 수 있다. 이것은 학습한 음성신호로부터 구해진 정보를 기초로 하여 학습하지 않은 음성신호에 대해서도 TDNN이 효과를 발휘한 것이라고 판단된다.

3. 유색잡음에 대한 효과

TDNN에 학습하지 않은 음성신호에 유색잡음을 부가한 경우의 입력신호 및 출력신호의 SD의 효과를 표 5에 나타낸다. 표 5의 F2에 유색잡음이 부가된 경우에 대해서 위너필터에 대해서는 최대 4.51(dB), 본 방식에 대해서는 최대 9.70(dB)이 각각 개선되었으며, 위너 필터에 의한 수법과 비교한 경우에도 학습하지 않은 음성에 대한 SD의 값은 모든 음성에 대해서 효과가 있었다.

학습에 의한 FFT 진폭성분의 변화의 예로서 "aioi" 제4프레임에 대한 학습신호(실선)와 입력신호(점선)의 비교를 그림 7에 나타내며, "aioi"를 학습하지 않은 시점에서의 학습신호(실선)와 출력신호(점선)의 비교를

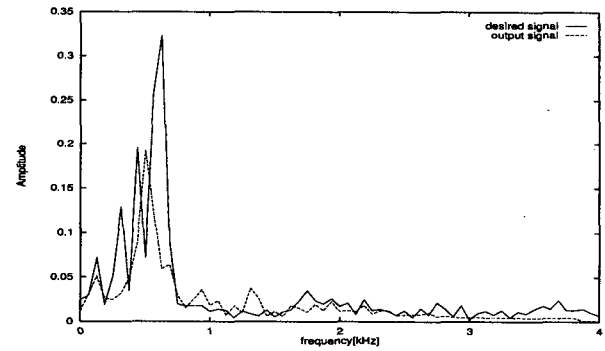


그림 8. 학습 및 출력신호의 FFT진폭성분의 비교 (학습 전)
Fig. 8. Comparison of FFT amplitude component for training signal and output signal(before trained).

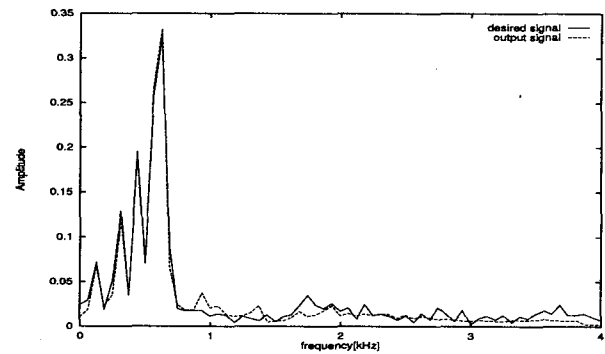


그림 9. 학습 및 출력신호의 FFT진폭성분의 비교(학습 후)
Fig. 9. Comparison of FFT amplitude component for training signal and output signal(after trained).

그림 8에 나타낸다. 그림 9는 "aioi"를 학습한 시점에서의 학습신호(실선)와 출력신호(점선)의 비교를 나타낸다. 그림들을 비교함으로써 "aioi"를 학습함으로써 FFT 진폭성분이 충분히 잡음이 없는 상태에 접근한 것을 알 수 있다.

이상의 결과로부터 학습 시에 사용한 음성신호뿐만 아니라 학습하지 않은 음성신호에 대해서도 화자에 관계없이 본 방식에 의한 잡음제거가 효과적인 것을 말할 수 있다. 특히 입력 SNR이 -10(dB) 정도의 극히 열악한 조건에서도 잡음 제거 효과가 높은 것을 확인하였다.

VI. 결 론

본 논문에서는 TDNN을 사용하여 잡음을 제거하는 시스템을 제안하여, 이것이 스펙트럴 왜곡률(SD)에서 유효하다는 것을 백색잡음 및 유색잡음에 대해서 실험

적으로 증명하였다.

본 알고리즘을 사용하여 음성신호에 포함되는 백색 잡음 및 유색잡음을 제거하는 실험을 SNR이 약 -10(dB)의 환경 하에서 실시하였다. 이 결과, 백색잡음을 제거하는 실험에 대해서는 TDNN에 학습시키지 않았음에도 불구하고 본 논문에서 사용한 학습시킨 음성신호 및 학습하지 않은 음성신호에 대해서 잡음제거의 효과가 구해졌다. 더욱이 위너필터에 의한 방법과 비교하여도 본 방법이 유효하다는 것을 확인할 수 있었다. 또한, 유색잡음을 제거하는 실험에 대해서도 본 논문에서 사용한 모든 음성신호에 대해서도 잡음제거의 효과를 확인할 수 있었다. 결론적으로 본 연구에서는 입력 SNR이 -10(dB) 정도의 극히 열악한 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 TDNN에 의한 본 방식이 백색잡음 및 유색잡음에 대해서 효과적이라는 것을 실험적으로 확인하였지만, 향후의 연구과제로서는 다양한 유색잡음에 의해서 열화된 음성에 대해서도 더욱 강화하는 방법의 검토가 필요하다고 생각된다. 향후, 이러한 검토를 더욱 심도있게 하여 본 방법을 발전시킴으로써 보다 효과적인 응용법이 전개될 것이라고 생각하고 있다.

참 고 문 헌

- [1] K. K. Paliwal, "Neural net classifiers for robust speech recognition under noisy environments", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 429-432, April 1990.
- [2] J. T. Chien, L. M. Lee, and H. C. Wang, "Noisy speech recognition by using variance adapted hidden Markov models", IEE Electronics Letters, Vol. 31, No. 18, pp. 1555- 1556, 1995.
- [3] W. G. Knecht, M. E. Schenkel, and G. S. Moschytz, "Neural network filters for speech enhancement", IEEE Trans. Speech and Audio Processing, Vol. 3, No. 6, pp. 433-438, 1995.
- [4] 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.
- [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] T. V. Sreenivas and P. Kimapure, "Codebook constrained wiener filtering for speech enhancement", IEEE Trans. Speech and Audio Processing. Vol. 4, No. 5, pp. 383-389, 1996.
- [8] S. Oh, V. Viswanathan, and P. Papamichalis, "Hands-free voice communication in an automobile with a microphone array", IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 92, No. 1, pp. 281-284, 1992.
- [9] B. Widrow, et al., "Adaptive noise cancelling: Principles and applications", Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, 1975.
- [10] M. R. Sambur, "Adaptive noise cancelling for speech signals", IEEE Trans. Acoust., Speech, Signal Processing. Vol. 26, No. 5, pp. 419-423, 1978.
- [11] A. Ishida and H. Gobata, "Speech/Non-speech Discrimination under Real Life Environments". J. Acoust. Soc. Japan, Vol. 47, No. 12, pp. 911-917, 1991.
- [12] 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.
- [13] M. Miyatake, H. Sawai, and K. Shikano, "Training Methods and Their Effects for Spotting Japanese Phenomes Using Time-Delay Neural Networks", IEICE, Vol. J73-D-II, No.5, pp. 699-706, 1990.
- [14] K. Itoh, N. Kitawaki, and K. Kakehi, "A Study of Objective Quality Measures for Digital Speech Waveform Coding Systems", IEICE, Vol. J66-A, No. 3, pp. 274-281, 1983.

— 저 자 소 개 —



최 재 승(정회원)

1989년 조선대학교 전자공학과 졸업(공학사)

1995년 일본 오사카시립대학 정보통신공학과(공학석사)

1999년 일본 오사카시립대학 정보통신공학과(공학박사)

2000년 일본 마쓰시타 전기산업주식회사 AVC사 연구원

2002년~현재 경북대학교 디지털기술연구소 연구원, 프로젝트 리더

<주관심분야: 음성신호처리, 잡음제거, 신경망, 디지털 TV 등>