

논문 2009-46SP-2-14

음성신호의 선형예측계수에 의한 잡음량의 인식

(Recognition of Noise Quantity by Linear Predictive Coefficient of Speech Signal)

최재승*

(Jae Seung Choi)

요약

잡음환경 하의 회화에서 잡음량을 줄이고 신호처리 시스템의 성능을 향상시키기 위해서는 잡음량에 따라서 적응적으로 처리되는 신호처리 시스템이 필요하다. 따라서 본 논문에서는 선형예측계수를 사용하여 잡음량을 인식하는 방법을 제안하며, 본 잡음량 인식은 다양한 배경잡음에 의하여 열화된 3종류의 음성이 신경회로망에 의하여 학습되어진다. 제안한 잡음량 인식의 성능은 다양한 잡음에 대하여 인식율을 사용하여 측정되었다. 본 실험에서는 Aurora2 데이터베이스를 사용하여 여러 잡음에 대하여 평균적으로 약 98.4% 이상의 양호한 인식결과를 확인할 수 있었다.

Abstract

In order to reduce the noise quantity in a conversation under the noisy environment, it is necessary for the signal processing system to process adaptively according to the noise quantity in order to enhance the performance. Therefore this paper presents a recognition method for noise quantity by linear predictive coefficient using a three layered neural network, which is trained using three kinds of speech that is degraded by various background noises. The performance of the proposed method for the noise quantity was evaluated based on the recognition rates for various noises. In the experiment, the average values of the recognition results were 98.4% or more for such noise using Aurora2 database.

Keywords : Linear predictive coefficient, recognition rate, noise quantity, neural network

I. 서론

근년 음성인식기술은 음성정보처리기술의 발달과 더불어 다양한 분야에서 실용화가 진행되고 있으며, 이에 대한 연구가 활발히 진행되고 있다. 그러나 음성인식 기술을 상업적으로 적용하기 위해서는 여러 가지 기술적인 문제를 해결해야 한다. 이러한 문제들 중에서 가장 중요한 요소는 음성에 추가되는 배경잡음의 영향을 줄이는 일이다^[1~4]. 이러한 배경잡음은 인식 대상 음성에 추가되어 음성인식 성능을 크게 저하시키게 된다.

최근에 신경회로망(Neural Network, NN)은 식별에

있어서 상당히 효과적인 능력이 있으며, 음성 및 문자의 식별에 대해서도 많은 성과를 올리고 있다^[4~7]. 또한 음성 중에서 잡음을 경감하기 위해서는 잡음의 강도에 따라서 각각 적당한 처리를 할 필요가 있다. 즉, 잡음의 크기를 인식하는 것이 상당히 중요하다. 참고문헌 [5]에서는 음성에 포함되는 잡음의 크기를 신경회로망을 사용하여 학습하였으며, 이 때 입력되는 음성데이터의 각 프레임의 일부분의 데이터들을 사용하여 학습한 결과를 나타내었다. 또한 신경회로망의 입력데이터로는 캡스트럼 변환(cepstral transform)된 저역부의 캡스트럼 데이터를 신경회로망에 입력하여 잡음량을 추정하였다. 본 논문에서는 배경잡음의 영향을 줄여서 음성인식 시스템의 성능을 향상시키고 다양한 음성인식기의 입력으로 사용하기 위하여, 선형예측분석에 의한 선형예측계수를

* 정희원, 신라대학교 전자공학과
(Department of Electronics Engineering,
Silla University)

접수일자: 2008년8월1일, 수정완료일: 2009년2월27일

신경회로망의 입력으로 한 시스템을 구축하고자 한다. 본 논문에서 사용하는 신경회로망의 입력데이터로는 각각의 프레임의 데이터를 사용하여 학습시키며, 신경회로망의 학습조건 및 학습방법 등을 바꾸어 음성 중의 잡음량을 인식하여 이러한 잡음을 경감하는 것을 목적으로 한 연구를 진행한다.

본 연구의 목적을 달성하기 위하여 본 논문에서는 잡음과 음성신호의 특징을 가진 선형예측계수(Linear Predictive Coefficient, LPC)를 신경회로망의 입력으로 하여 3종류의 잡음량을 인식하는 방법을 제안한다.

II. 음성신호의 선형예측분석

음성신호의 표본값 사이에는 커다란 상관관계가 있으며 음성의 특징 추출을 위하여 이것을 이용한 예측부호화가 실시되어진다^[8]. 이러한 예측의 개념을 일반화하여 다음 식과 같이 음성파형의 연속된 $p+1$ 개의 표본값 사이에 높은 선형예측성이 있다고 가정한다.

$$\hat{x}_n = a_1x_{n-1} + a_2x_{n-2} + \dots + a_px_{n-p} + e_n \quad (1)$$

따라서 음성신호의 현재 값은 p 개의 과거의 값 $x_{n-1}, x_{n-2}, \dots, x_{n-p}$ 로부터 예측된다. 여기에서 $a_i(i=1, \dots, p)$ 는 선형예측계수이며, e_n 은 식 (2)와 같이 실제 입력된 값과 예측된 값과의 차이를 나타내는 선형예측오차이다.

$$e_n = x_n - \hat{x}_n \quad (2)$$

즉, 어떤 시점 n 에서의 표본값은 그 이전의 연속한 p 개의 표본값의 선형 하중값으로서 근사할 수 있다고 가정한다. 또한 음성파형은 정상적으로 선형예측계수는 시간적으로 변화하지 않는다고 가정한다. 이러한 가정하에서, 선형예측오차 e_n 의 2승 평균값을 최소로 하도록 선형예측계수를 결정한다. 식 (1)로부터 선형예측오차 e_n 의 2승 평균값을 계산하여 장시간 평균을 구한다. 이 때, 시간평균 조작은 선형한 조작 방법을 이용하여 평균적인 각항 별로 평균한다. 또한 선형예측계수는 시간 불변임으로 다음 식과 같이 구해진다.

$$\begin{aligned} \overline{e_n^2} = & v_{00} + a_1v_{01} + a_2v_{02} + \dots + a_pv_{0p} \\ & + a_1v_{10} + a_1^2v_{11} + a_1a_2v_{12} + \dots + a_1a_pv_{1p} \\ & \dots \dots \dots \\ & + a_pv_{p0} + a_pa_1v_{p1} + a_pa_2v_{p2} + \dots \\ & \dots + a_pa_{p-1}v_{pp-1} + a_p^2v_{pp} \end{aligned} \quad (3)$$

여기에서 $v_{ij} = \sum_n x_{n-i} \cdot x_{n-j}$ 으로 파형 $\{\hat{x}_n\}$ 의 상관계수이다. 위의 식을 최소로 하는 $\{a_i\}$ 의 조건으로는, 각 a_i 에 의한 편미분을 0으로 했을 때 신호가 정상임으로, $v_{ij} = v_{i-j}$ 를 사용하여 다음의 p 개의 식이 구해진다.

$$\begin{aligned} 2(a_1v_0 + a_2v_1 + \dots + a_pv_{p-1} + v_1) &= 0 \\ 2(a_1v_1 + a_2v_0 + \dots + a_pv_{p-2} + v_2) &= 0 \\ \dots \dots \dots \\ 2(a_1v_{p-1} + a_2v_{p-2} + \dots + a_pv_0 + v_p) &= 0 \end{aligned} \quad (4)$$

구하는 해는 이 p 의 식을 연립하는 것이며, 이 결과에 의하여 극치가 구해져, 이것이 유일한 해이면 구하는 최소치이다. 위의 식을 정리하면 다음식과 같은 연립 p 원 1차방정식이 된다.

$$\begin{pmatrix} v_0 & v_1 & v_2 & \dots & v_{p-1} \\ v_1 & v_0 & v_1 & \dots & v_{p-2} \\ v_2 & v_1 & v_0 & \dots & v_{p-3} \\ \dots & \dots & \dots & \dots & \dots \\ v_{p-1} & v_{p-2} & v_{p-3} & \dots & v_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ p \end{pmatrix} = - \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ \vdots \\ v_p \end{pmatrix} \quad (5)$$

이와 같이 하여 구해진 선형예측계수는 분석의 대상인 일련의 데이터를 전극 모델에 의하여 생성하였을 때의 시스템의 요소가 된다. 따라서 이 계수로서 생성 모델의 정보가 추출되어, 이것들을 부호화함으로서 고능률 부호화가 가능하다.

III. 실험 데이터 및 신경회로망

1. 음성 및 잡음 데이터

본 실험에서 사용한 음성 데이터는 8 kHz의 샘플링 주파수를 가진 환경에서 녹음된 연결된 영어숫자로 구성된 Aurora2 데이터베이스^[9]이다. 제안한 시스템은 Aurora2 데이터베이스로부터의 테스트 셋 A, B, C의 음성데이터와 테스트 셋 A의 자동차(car noise), 지하철잡음(subway noise), 테스트 셋 C의 도로잡음(street noise) 그리고 컴퓨터에 의해서 작성된 가우스 백색잡음(white noise) 등의 배경잡음을 사용하여 평가하였다. Aurora2 데이터베이스의 각 테스트 셋에는 백색잡음을 제외한 이러한 잡음들이 음성데이터에 혼입된 잡음이 중첩된 데이터(즉, SNR_{in}(Input Signal-to-Noise Ratio)=∞, 15 dB, 5 dB)들이 포함되어 있다. 백색잡음에 대해서는 Aurora2 데이터베이스의 각 음성데이터에 별도로 백색잡음을 부가하여 실험 데이터를 구하였다. 따라서 본 실험에서는 3종류의 입력 신호대잡음비

(SNR_{in}=∞, 15 dB, 5 dB)와 같이 잡음이 부가된 음성 신호를 사용하여 신경회로망을 학습시켰다. Aurora2 데이터베이스를 사용할 경우에 백색잡음, 자동차잡음, 지하철잡음을 Aurora2 데이터베이스의 음성신호에 부가한 후에 신경회로망이 학습되었다.

2. 신경회로망의 구성

본 논문에서 잡음량의 인식방법으로 제안한 신경회로망의 학습법은 Rumelhart^[10]에 의해서 제안된 2층 오차역전파의 학습을 다층 네트워크 전체의 학습에 확장시킨 방법으로 오차역전파학습법으로 불리운다. 이 학습법은 입력값이 주어졌을 때 교사신호와 출력값의 오차를 최소로 하며 신경세포 사이의 결합계수를 조절하는 방법이다. 그림 1은 본 실험에 사용한 3층 구조의 퍼셉트론(perceptron)형^[11]의 신경회로망의 구조이며, 입력층의 10 유닛, 중간층의 15 유닛, 출력층의 3 유닛으로 구성된 네트워크를 사용한다.

m층의 네트워크로 생각하여, k층의 제i유닛의 입력의 총합을 i_i^k , 출력을 o_i^k 로 하여, k-1층의 제i유닛으로부터 k층의 제j유닛에의 결합의 하중을 w_{ji}^{k-1k} 로 한다. 각 유닛의 입출력관계를 부여하는 함수를 f로 하며, 이러한 변수 간의 관계는 다음 식과 같다.

$$i_i^k = \sum_j w_{ji}^{k-1k} o_j^{k-1} - \theta_i^k \tag{6}$$

$$o_i^k = f(i_i^k) \tag{7}$$

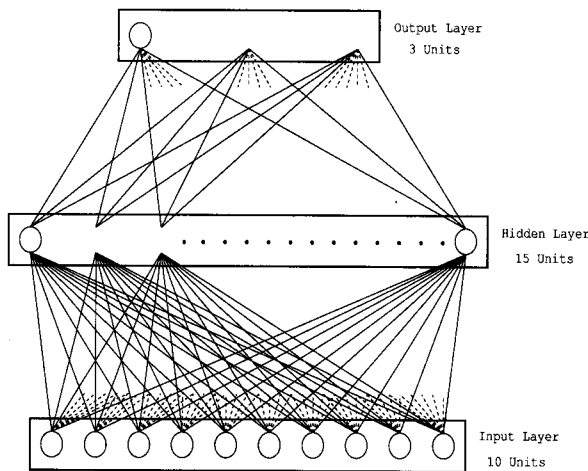


그림 1. 본 시스템에 사용한 3층 구조의 신경회로망 구성

Fig. 1. Construction of three layer neural network system.

여기에서 θ_i^k 는 제k층 제i유닛의 문턱치(threshold)를 나타내며, 각 유닛의 입력 및 출력관계는 식 (8)과 같은 비선형 함수를 사용한다.

$$f(x) = \left(\frac{1}{1 + \exp(-x)} \times 2 \right) - 1 \tag{8}$$

본 실험에서는 각 프레임에서 잡음이 없는 음성(SNR_{in}=∞), 잡음이 적은 음성(SNR_{in}=15dB), 잡음이 많은 음성(SNR_{in}=5dB)의 3종류를 인식할 수 있도록 신경회로망의 출력층의 유닛수를 3으로 하여 학습시킨다. II장의 음성신호의 선형예측 분석에 의하여 구해진 10차의 선형예측계수가 입력층의 각 유닛에 입력되며, 신경회로망의 교사신호는 (T1) SNR_{in}=∞일 때 (1.0, -1.0, -1.0), (T2) SNR_{in}=15dB일 때 (-1.0, 1.0, -1.0), (T3) SNR_{in}=5dB일 때 (-1.0, -1.0, 1.0)와 같이 3종류의 목표치가 주어진다. 즉, 각 프레임에서 잡음량이 (T1) SNR_{in}=∞, (T2) SNR_{in}=15dB, (T3) SNR_{in}=5dB로 출력되도록 잡음량을 인식한다. 그리고 각각의 유닛의 하중 w_{ij} 는 -0.005~0.025의 범위 내의 랜덤한 값으로 초기화하며, 학습계수 $\alpha = 0.2$, 가속도계수 $\beta = 0.5$ 로 한다. 또한 본 실험에서는 최대 학습횟수를 10,000회로 하여, 각 음성데이터에 있어서 결합하중 w_{ij} 의 초기값을 바꾸어서 5회씩 시행한다.

그림 2는 백색잡음, 자동차잡음, 지하철잡음을 Aurora2 데이터베이스의 음성신호에 부가한 경우의 10-15-3 네트워크에 대한 신경회로망의 학습 시의 오차곡선을 나타낸다. 여기에서 오차곡선은 하나의 네트워크에 대해서 5회의 시행에 대한 평균값이며, 그림에서와 같이 학습횟수가 약 8000회를 초과하면 오차의 수렴이 거의 0에 가까워지는 것을 알 수 있다.

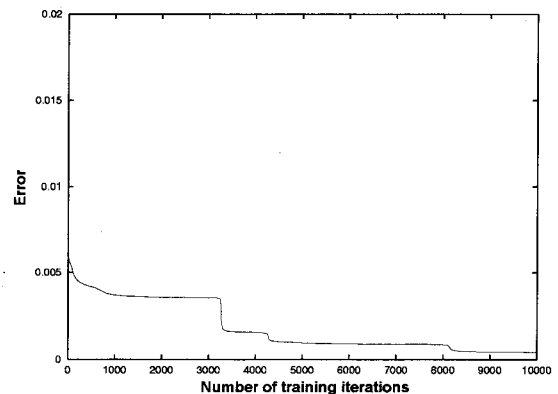


그림 2. 학습 시의 오차 곡선
Fig. 2. Error curves of training.

IV. 선형예측계수에 의한 잡음량 인식

본장에서는 3층 구조의 퍼셉트론형의 신경회로망을 사용하여, 선형예측계수를 입력으로 하여 각 프레임에서 이 잡음량을 인식하는 것을 목적으로 한 실험에 대하여 기술한다.

1. 제안한 시스템 구성

II장에서 기술한 방법으로, 음성신호의 표본값을 선형예측 분석하여 각 프레임 별로 10차의 선형예측계수를 구하여 신경회로망의 입력으로 한다.

본 연구에서 제안한 잡음량 인식 시스템을 그림 3에 나타낸다. 표본 주파수 8 kHz의 이산 시간신호 $x(t)$ 를 해밍창 $w_1(t)$ 에 의해 256 표본의 프레임으로 분리하여 각 프레임의 실효치 R_f 를 구하여 이 값이 $R_m/3$ 보다 적은 프레임만을 사용한다. 여기에서 R_m 은 각 문장의 문장 전체에서 구한 실효치이다(식 (9), (10), (11) 참조). 각 프레임의 표본값을 선형예측 분석하여 10차의 선형예측계수를 구한다. 이렇게 함으로써 원래의 표본값은 10차의 선형예측계수와 잔차신호로 완전히 복구가 가능하다. 분석하고자 하는 음성신호의 표본값에 잡음이 중첩된 경우에 잡음은 선형예측계수와 잔차신호의 모두에 영향을 미치지만 본 실험에서는 잔차신호에는 조작을 하지 않고 선형예측계수만을 사용하여 이 값의 차이로부터 잡음량의 인식을 한다. 그리고 10차의 선형예측계수를 신경회로망의 입력으로 사용하여 학습을 한다. 신경회로망의 학습을 통해 얻어진 가중치를 저장한 후, 학습에 사용되지 않은 잡음이 중첩된 음성데이터의 선형예측 계수를 신경회로망의 입력으로 받아 III.2절에서 기술한 교사신호 T1, T2, T3의 목표치와 비교하여 각 프레임에서 잡음량을 인식한다.

위에서 기술한 바와 같이 본 실험에서는 식 (9)의 각 음성데이터의 평균전력 R_m 을 구하여 식 (10)의 각 프레임에서의 평균전력 R_f 와의 관계가 식 (11)과 같이 되도록 프레임의 선형예측계수 만을 사용한다.

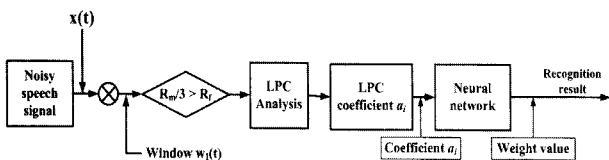


그림 3. 잡음량 인식 시스템

Fig. 3. Block diagram of the proposed system.

$$R_m = \sqrt{\frac{\sum_{i=1}^M S_i^2}{M}} \tag{9}$$

$$R_f = \sqrt{\frac{\sum_{i=1}^{256} S_{i \cdot f}^2}{256}} \tag{10}$$

$$\frac{R_m}{3} > R_f \tag{11}$$

단, M은 전샘플수, S_i 는 문장 전체에서 구한 잡음이 중첩된 음성신호의 표본값, $S_{i \cdot f}$ 는 각 프레임 내의 잡음이 중첩된 음성신호를 나타낸다.

2. 학습 문장 간의 거리 측정

학습 문장 사이에서의 패턴의 유사성을 측정하기 위해서 아래의 식과 같이 정의된 선형예측계수의 거리 D를 측정하여, 거리와 신경회로망에 의한 잡음량의 인식율과의 관계를 명확하게 하였다.

$$D_{i \cdot j} = \sqrt{\sum_{k=1}^{10} (I_k - J_k)^2} \tag{12}$$

여기에서, $i \cdot j$ 는 비교하는 잡음의 크기를 나타낸다. 예를 들면, T2(SNR_{in}=15dB)와 T3(SNR_{in}=5dB)의 비교라면, $D_{T2 \cdot T3}$ 가 된다. I_k, J_k 는 각각 비교하는 프레임의 선형예측계수의 값이다.

표 1은 선형예측계수 방식에 의한 거리측정결과와 참고문헌 [5]에서 사용한 켈스트럼계수 방식에 의한 거리측정결과를 나타내고 있으며, 학습에 사용된 10개의 네트워크에 대한 평균값의 결과이다. 켈스트럼계수 방식은 입력음성을 해밍창을 통과 시킨 후에 켈스트럼 변환하여 저역에 해당하는 10개의 켈스트럼계수를 구하는 방식이다.

표의 결과로부터, 1프레임 중의 선형예측계수의 거리 D의 값이 SNR_{in}=15dB(T2)과 SNR_{in}=5dB(T3)의 경우($D_{T2 \cdot T3}$), 다른 경우와 비교하면 거리가 극히 밀접하

표 1. 선형예측계수 및 켈스트럼 거리의 비교

Table 1. Comparisons for distance of linear predictive coefficient (LPC) and cepstral coefficient.

	$D_{T1 \cdot T2}$	$D_{T1 \cdot T3}$	$D_{T2 \cdot T3}$
LPC	0.359	0.396	0.049
cepstral	0.635	0.714	0.221

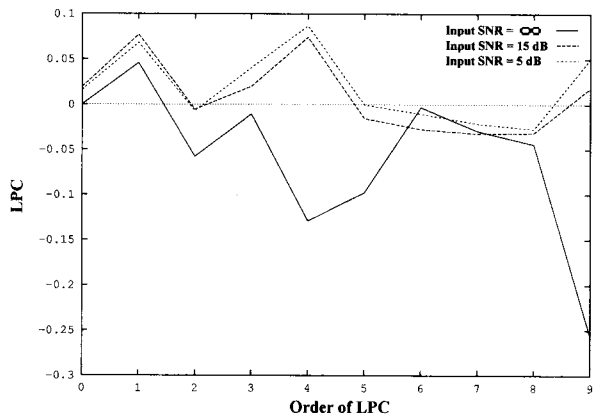


그림 4. 잡음량차이에 의한 선형예측계수의 예(프레임=10)

Fig. 4. An example of LPC by the difference of noise quantity (in case of 10 frame).

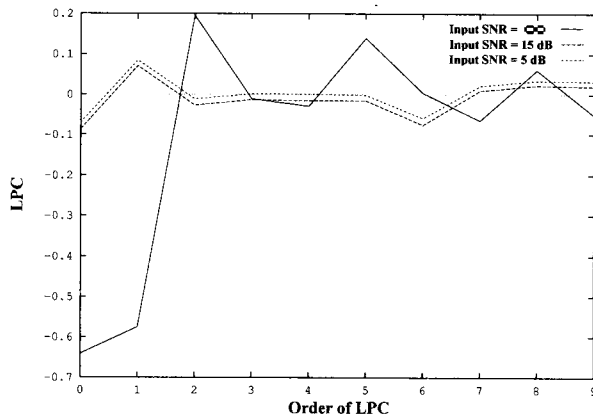


그림 5. 잡음량차이에 의한 선형예측계수의 예 (프레임=100)

Fig. 5. An example of LPC by the difference of noise quantity (in case of 100 frame).

기 때문에, 신경회로망에 의한 인식이 쉽지 않을 거라고 추측되어진다. 또한 캡스트럼거리 방식이 본 논문에서 제안한 선형예측계수 방식보다 거리가 상당히 떨어져 있으므로, 각 데이터간의 중첩이 적고 신경회로망에 의한 인식이 본 방식보다 어느 정도 용이하게 가능하다는 것을 판단할 수 있다. 이러한 방식들에 의한 인식 결과를 제V장에서 자세히 나타낸다.

그림 4와 그림 5는 입력으로 하는 선형예측계수의 예를 그래프로 나타낸 것으로(제10프레임과 제100프레임의 선형예측계수), 본 실험에서 사용한 프레임의 선형예측계수 뿐만 아니라 거의 전 프레임에 대해서, $SNR_{in}=15dB(T2)$ 과 $SNR_{in}=5dB(T3)$ 에 대하여 선형예측계수에는 거의 차이 없었다. 그림에서 알 수 있듯이, 선형예측계수의 간격이 상당히 접근해 있음에도 불구하고

제V장에서 나타내는 인식을 실험과 같이 각 학습데이터를 양호하게 인식할 수 있었다.

V. 실험 결과

본 논문에서는 출력되어진 학습결과와 학습신호를 비교하여 프레임에서의 잡음량의 인식율을 구한다. 잡음량 인식율 P는 다음 식과 같이 정의한다.

$$P = \frac{F_c}{F_s} \times 100 \tag{13}$$

여기에서, F_c 는 잡음량이 정확하게 인식된 프레임수를 나타내고, F_s 는 식 (11)을 만족하는 입력에 사용된 총 프레임수를 나타낸다.

표 2와 표 3은 Aurora 2 데이터베이스의 테스트 셋 C로부터 임의적으로 20개의 문장을 선택하여, 각 음성으로부터 구한 선형예측계수를 신경회로망의 입력으로 하여 실험을 실시한 학습 결과를 각 잡음에 대하여 나타낸 평균이다. 표 2는 본 논문에서 제안한 선형예측계수에 의한 잡음량의 인식율을 나타내며, 표 3은 비교를 위하여 캡스트럼계수 방식에 의한 인식율을 나타낸다.

표 2의 선형예측계수에 의한 각 잡음에 대한 인식율의 평균값과 표 3의 캡스트럼계수에 의한 각 잡음에 대한 인식율의 평균값을 비교하면, 본 논문에서 제안하는

표 2. 선형예측계수에 의한 각 잡음에 대한 인식율
Table 2. Recognition rates for each noise by LPC.

Type of noise	Recognition rates (%)		
	T1 (∞)	T2 (15 dB)	T3 (5 dB)
white	100.0%	99.4%	100.0%
Car	99.9%	98.2%	100.0%
Subway	99.5%	97.3%	99.8%
Street	97.9%	94.1%	95.5%
Average	99.3%	97.2%	98.8%

표 3. 캡스트럼계수에 의한 각 잡음에 대한 인식율
Table 3. Recognition rates for each noise by cepstral coefficient.

Type of noise	Recognition rates (%)		
	T1 (∞)	T2 (15 dB)	T3 (5 dB)
white	100.0%	99.1%	100.0%
Car	99.9%	97.8%	100.0%
Subway	99.0%	97.1%	99.5%
Street	97.6%	93.7%	94.9%
Average	99.1%	96.9%	98.6%

선형예측계수에 의한 잡음량 인식율이 캡스트럼계수에 의한 잡음량 인식율보다 양호하다는 것을 알 수 있으며, 또한 여러 잡음에 대하여 평균적으로 약 98.4% 이상의 높은 인식결과를 확인할 수 있었다. 또한 본 논문에서 제안한 표 2의 선형예측계수에 의한 3 패턴의 학습신호에 의한 학습결과로부터, $SNR_{in}=15dB(T2)$ 의 인식율이 다른 입력($T1$ 및 $T3$)보다 약간 인식율이 떨어지는 반면에, $SNR_{in}=\infty(T1)$ 과 $SNR_{in}=5dB(T3)$ 에서 상당히 좋은 인식결과를 볼 수 있었다.

다른 방식의 인식시스템과 비교하기 위하여 참고문헌 [12]와 참고문헌 [13]을 제시한다. 참고문헌 [12]에서는 20 kHz로 샘플링한 음성데이터를 25.6ms의 Hamming 창을 씌웠으며 창의 이동간격은 12.8ms로 하였다. 선형예측 캡스트럼 계수(Linear Predictive Cepstral Coefficients: LPC)의 차수는 28차로 하였다. 모음 인식실험은 잡음을 혼입하지 않은 경우(clean)와 Gaussian 백색잡음을 각각 SNR 20dB, 10dB, 5dB로 혼입하여 화자독립일 경우의 인식율은 각각 86.4%, 82.4%, 78.9%, 77.9%이었다. 참고문헌 [13]에서는 음성 신호의 샘플링 주파수를 12 kHz, 한 프레임을 256샘플, 프레임 길이를 21.33ms, 프레임 이동을 5ms하였으며, LPC 캡스트럼의 인식율은 SNR clean, 30dB, 25dB, 20dB, 15dB에 대하여 각각 50단어음성데이터에 대하여 89.28%, 88.78%, 88.44%, 86.93%, 86.16%인 결과를 보였다. 따라서 본 논문에서 제안하는 방법과 비교하면 본 방식이 약 10%~17% 정도 잡음량 인식율이 높은 것을 알 수 있다. 따라서 다른 방식과 비교하여도 본 방식의 효율성이 높다고 볼 수 있다.

이상의 결과로부터, 제IV.2절의 표 1의 거리측정 결과에서 알 수 있듯이 선형예측계수의 거리가 캡스트럼 계수의 거리보다 상당히 밀접해 있음에도 불구하고 선형예측계수에 의한 각 프레임에서의 잡음량 인식율이 높다는 것은 본 논문에서 제안한 방식이 상당히 유효하다는 것을 말할 수 있다.

VI. 결 론

본 논문에서는 신경회로망에 의한 3종류의 음성신호의 잡음량을 인식하는 것을 목적으로 하여, 선형예측계수를 입력으로 한 잡음량 인식의 실험을 실시하였다. 이 결과, 신경회로망의 파라미터 및 학습횟수 및 입력 데이터의 균형에 따라서 본 논문에서 추구하는 목적을

달성할 수 있었다. 그러나 음성인식 및 음성강조에 응용하기 위해서는, 다음과 같은 내용이 향후의 과제 및 목표라고 생각하며, 추후 더 상세한 연구를 실시하고자 한다. (1) 학습데이터에 정확하고 신속하게 수속하기 위한 중간층의 층수 및 유닛수의 선택 (2) 본 실험에서는, 식(11)의 조건에 적합한 프레임 만을 사용하였지만, 역으로 이 조건에 적합하지 않은 프레임을 사용하여 실험한다.

이상과 같이 다양한 잡음이 중첩된 음성신호에 대한 잡음량의 인식을 신경회로망을 통하여 실험적으로 확인하여 본 연구가 음성인식 및 음성신호처리에 효과적으로 응용될 것이라고 생각한다.

참 고 문 헌

- [1] S. V. Vaseghi, B. P. Milner: Speech Recognition in Impulsive Noise. International Conference on Acoustics, Speech, and Signal Processing, pp. 437 - 440, vol. 1, 1995.
- [2] O. Ichikawa, T. Fukuda, M. Nishimura: Local peak enhancement combined with noise reduction algorithms for robust automatic speech recognition in automobiles. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4869 - 4872, 2008.
- [3] 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.
- [4] 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.
- [5] 최재승, "음성강조에의 응용을 위한 신경회로망에 의한 잡음량 추정법", 전자공학회 논문지 제42권 SP편 제3호, pp. 129-136, 2005.
- [6] 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.
- [7] 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.
- [8] P.B. Patil: Multilayered network for LPC based speech recognition. IEEE Transactions on Consumer Electronics, Vol. 44, No. 2, pp. 435 -

- 438, 1998.
- [9] 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.
- [10] D. Rumelhart, "Parallel Distributed Processing, vol. 1 and 2, MIT Press, Cambridge, MA, 1986.
- [11] S. K. Pal, S. Mitra, "Multilayer perceptron, fuzzy sets, and classification", IEEE Transaction on Neural Networks, vol. 3, no. 5, pp. 683-697, 1992.
- [12] 윤태성, 심재성, "인간의 청각특성을 이용한 잡음 혼입시의 음성인식에 관한 연구", 창원대학교 산업기술연구소 논문집, 제6호, pp. 115-124, 1992.
- [13] Xueying Zhang, Yueling Guo, Xuemei Hou, "A speech Recognition Method of Isolated Words Based on Modified LPC Cepstrum", IEEE International Conference on Granular Computing, pp. 481-484, 2007.

 저 자 소 개



최재승(정회원)

1989년 조선대학교 전자공학과
학사졸업

1995년 일본 오사카시립대학
정보통신공학과 석사졸업

1999년 일본 오사카시립대학
정보통신공학과 박사졸업

2000년~2001년 일본 마쓰시타 전기산업주식회사
AVC사 연구원

2002년~2007 경북대학교 디지털기술연구소
책임연구원

2007년~현재 신라대학교 전자공학과 교수
<주관심분야: 디지털통신, 음성신호 및 영상신호
처리, 신경회로망, 유전알고리즘, 디지털 TV 및
멀티미디어 등>