

Neural-HMM을 이용한 고립단어 인식

正會員 金 鈞 秀* 正會員 金 昌 錫*

Isolated-Word Recognition Using Neural Network
and Hidden Markov ModelsYeun Soo Kim*, Chang Seok Kim* *Regular Members*

要 約

본 논문에서는 HMM(Hidden Markov Models)에서 문제점이 되는 개인차에 의한 변동을 흡수하고, 적은 학습 데이터로서 인식률을 향상시키기 위하여 신경회로망을 이용한 NN-HMM(Neural Network-Hidden Markov Models)에 의해 한국어 인식에 관하여 연구하였다.

이 방법은 HMM과 신경회로망의 출력을 각각 독립적인 인식값으로 가정하여 두 시스템의 확률곱으로 서로 보정되어 최대 인식확률의 음성모델을 인식하는 음성인식 시스템이다. 본 방법의 타당성을 평가하기 위하여 남, 여화자가 28개의 DDD 지역명을 발성한 음성데이터로 실험한 결과, 이산분포 HMM에 의한 방법에서는 91[%], 신경회로망에 의한 방법에서는 89[%], 제안된 방법에서는 95[%]의 향상된 인식률을 얻으므로써 인식성능의 우수함을 확인하였다

ABSTRACT

In this paper, a Korean word recognition method which use Neural Network and Hidden Markov Models(HMM) is proposed to improve a recognition rate with a small amount of learning data. The method reduces the fluctuation due to personal differences which is a problem to a HMM recognition system.

In this method, effective recognizer is designed by the complement of each recognition result of the Hidden Markov Models(HMM) and Neural Network. In order to evaluate this model, word recognition experiment is carried out for 28 cities which is DDD area names uttered by two male and a female in twenties. As a result of testing HMM with 8 state, codeword is 64, the recognition rate 91[%], as a result of testing Neural network(NN) with 64 codeword the recognition rate is 89 [%]. Finally, as a result of testing NN-HMM with 64 codeword which the best condition in former tests, the recognition rate is 95[%].

*明知大學校 電子工學科
Department of Electronic Engineering,
Myong-Ji University
論文番號 : 92 - 119 (接受1992. 4. 7)

I. 서 론

음성은 인간 사이의 중요한 통신 수단이지만 인간과 기계 사이의 주요통신 수단은 되지 못했다. 그 이유는 아직까지 음성 인식 기술이 자유롭게 기계와 사람 사이의 통신을 할 수 있을 정도로 발전되지 못하였기 때문이다. 음성 인식이란 시스템에 입력된 음성을 정확히 인식하여 문자로 바꾸어 주거나 혹은 입력된 음성을 이해하여 적절히 요구에 대응하여 주는 것을 말한다. 음성을 정확히 인식하여 문자로 변환시켜주는 일은 주로 제한되고 고립된 단어 인식 시스템에서 처리되며 음성을 이해하는 일은 주로 대용량 연속 단어 인식 시스템에서 수행된다.

음성 인식 방법은 인식 방법에 따라 정합(pattern matching)에 의한 방법과 통계적인 방법, 신경회로망에 의한 방법으로 크게 구분 된다. 이러한 알고리즘에는 그 동안 DWA(Dynamic Time Warping), VQ(vector Quantization), HMM(Hidden Markov Model), TDNN(Time Delay Neural Network)^[1]등이 이용되었는데 그중에서도 HMM이 다른 방법보다 우수한 장점이 있는 관계로 많은 연구가 이루어지고 있다. HMM은 음성을 통계적으로 모델화하는 가장 유력한 방법이고, 이상적인 심분에 대해서 출력 확률이 정의된다. 그러나 이 HMM도 높은 인식률을 얻기 위해서는 다량의 학습 데이터가 필요하고, 개인차의 흡수에도 문제가 지적되고 있다.

따라서 본 연구에서는 소량의 학습 데이터로써 개인차의 흡수를 해결하기 위하여 HMM과 신경 회로망을 결합한 하나의 음성인식 시스템으로한 NN-HMM을 이용하는 두가지 독립적인 인식시스템의 상호 오인식을 확률 보정하여 줌으로써 좋은 결과를 얻을수 있었다. 본 연구에서는 제안된 NN-HMM을 이용한 음성 인식 방법 이외에도 비교를 위하여 이산 분포HMM(DHMM)에 의한 방법, 신경회로망(Neural Network : NN)을 이용한 방법과 비교 평가하였다. 여기에서 사용한 음성 인식 대상어로는 28개의 DDD 지역명을 대상으로 남성 2인, 여성 1인의 화자가 각각 5번 발성한 음성데이터로하였고 벡터-양자화 코드북을 작성하는데 사용된 벡터(vector)는 14차 선형예측계수(L: near Predictive Coefficient : LPC)로 하였다.

1. 벡터양자화 코드북작성

벡터양자화의 가장 큰 목적은 데이터의 압축으로 이에 따른 왜곡을 되도록 적게하여, 최대 한도로 bit rate를 줄이는데 있다. 그런데 1948년의 Shannon에 의해 데이터 압축과 관련한 정보이론에서 스칼라 보다 벡터를 부호화 하면, 보다 효율적이라는 연구가 보고 되고, 1975년 Lloyd에 의해 반복과정에 의한 벡터양자화 방법이 발표된 이후, 신호를 개별적인 값으로 보는 것이 아니라 몇 개의 신호인 벡터로 묶으로써, 벡터의 성분 보다는 벡터 자체를 양자화하는 벡터양자화의 연구가 시작되었다. 이 과정은 통신에서 데이터압축을 위해 사용되고 있지만, 음성처리의 인식 방법으로 사용되고 있으며, 1985년 Rabiner는 벡터양자화를 DTW와 관련된 전처리 과정으로서 계산량을 줄이면서, 인식률에 기여하는 방법으로 사용하였다.^[2] 본 연구에서도 음성신호의 파라미터인 관측시계열을 표현하는 방법으로 사용하였다.

본 연구에서는 전 단어를 학습 데이터로 하여 남성 화자 2명과 여성 화자 1명의 화자가 3번째 발성한 데이터로 하나의 코드북을 만들었다. 이때 코드북을 구하는 방법으로는 clustering기법중 평균왜곡이 최소가 되도록하는 방법으로 반복집단화 알고리즘(iterative clustering algorithm)인 K-means 알고리즘^[3]을 사용하였다. 그런데, K-means 알고리즘은 국부최적 알고리즘(local optimal algorithm)이므로, 그 성능은 초기중심벡터에 의해 영향을 받는다. 초기중심벡터를 잡는 방법은 계산시간과 성능면에서 가장 효과적인 방법으로는 분할(splitting)방법인 LBG 집단화 알고리즘^[4]을 사용하였다. 벡터양자화에서 사용하는 선형예측계수를 이용한 거리적도에는, Itakura-Saito의 log-likelihood ratio(LLR) 거리측정이 사용되었고, 이들은 다음과 같이 주어진다.

$$d(x,y) = \frac{a_y^T R_x a_y}{a_x^T R_x a_x} - 1.0 \quad (1)$$

여기서, $d(x,y)$ 는 비교 데이터 x 와 y 의 거리이고, a_x 는 미지벡터 x 의 선형예측계수벡터이고 a_y 는 미지벡터 y 의 선형예측계수벡터이다. 또 a_x^T 는 a_x 의 전치행렬이고, R_x 는 코드벡터의 자기상관계수벡터이다.

식(1)를 이용하면 다음과 같은 LLR 거리 d_{LLR} 를 얻을 수 있다.

$$d_{LLR} = \log \left[\sum_{i=0}^N V_i(i) R_a(i) \right] \quad (2)$$

II. NN-HMM을 이용한 음성인식

여기에서, V_i 는 P차 선형예측계수 분석시 생기는 잔차에너지(residual energy)로 정규화한 자기상관 계수이고, $R_a(i)$ 는 선형예측계수벡터의 자기 상관벡터이다.

2. 음성 인식 알고리즘

2-1. HMM(Hidden Markov Model)

HMM 알고리즘은 훈련 과정 및 인식 과정을 수행하는 알고리즘으로서 높은 인식률과 빠른 인식 시간 때문에 대용량의 음성 인식 시스템에 많이 사용되고 있다. HMM 알고리즘의 기본적인 사상은 음성이 확률통계적인 Markov 모델로 모델링될 수 있다는 가정 하에 훈련 과정에서 Markov 모델의 파라미터를 얻어 기준 Markov 모델을 만들고 인식 과정에서는 입력 음성과 가장 유사한 기준 Markov 모델을 찾아냄으로써 인식한다. Markov 모델로서 Hidden Markov 모델을 사용하는데 그 이유는 음성 패턴의 다양한 변화를 수용하기 위해서이다. Hidden Markov 모델이란 이중 통계적인 처리(stochastic process)로서 상태 선정에 관한 통계적인 처리와 매 상태마다 음성 패턴이 발생될 출력 확률(output probability)에 관한 통계적인 처리로 구성된다. 즉 음성 패턴의 각 특징을 상태의 선정 확률과 출력 확률 등으로 표현하여 준다. 여기서 hidden이란 의미는 상태가 음성 패턴에 관계 없이 모델 속에 숨어 있다는 것을 말한다.

본 연구에서는 관측 심볼을 이산적 출력인 VQ 대표인덱스로 하는 이산분포 HMM(DHMM)^[5]을 이용하여 실험하였다. 실험에 사용된 알고리즘에는 전향-후향 알고리즘(forward-backward algorithm)인 Baum-Welch 알고리즘을 사용하여 학습하였으며, viterbi 알고리즘^[6]으로 인식을 수행하였다.

2-2. 신경회로망

신경회로망^{[7][8]}은 인간의 뇌세포를 간단히 모델링하고 모델된 뇌세포들을 연결시켜줌으로써 인간의 뇌가 하는 역할을 수행시켜 주는 알고리즘이다. 본 연구에서는 입력층, 중간층, 출력층으로 구성되고 각 층의 신호는 상층으로 향하는 feed-forward형 회로망이다. 각 유니트의 출력과 결합각도를 나타내는 웨이트들의 총합에 의해 결정되고, 이것은 비선형 단조증가 특성을 갖는 시그모이드함수(sigmoid function)에 의해 상층으로 출력된다.

$$x_j = \sum y_i W_{ji} \tag{3-1}$$

$$y_i = 1 / (1 + \exp(-x_i)) \tag{3-2}$$

단층 퍼셉트론이란 한 개의 뉴런을 모델링한 것이며 입력 데이터와 출력 데이터가 주어지면 가중(weighting) 값을 설정할 수 있다. 음성 인식일 경우가 가중값이 곧 기준 패턴이 된다. 그러나 단층 퍼셉트론으로는 입력 데이터를 분류할 수 없는 경우(exclusive OR 상태)도 있기 때문에 실제로 MLP가 음성 인식 시스템에 이용된다. 이때 MLP의 입력 데이터로는 음성 특징 추출된 데이터가 된다. 그런데 음성은 주파수에 의한 공간적인 특성과 발생 시간에 의한 시간적인 특징이 모두 포함되어 있기 때문에 MLP를 발생 시간이 긴 음성 인식에 이용하기 위해서는 시간적인 특징도 포함할 수 있어야 한다. 이러한 문제점을 해결하기 위한 MLP의 변형이 TDNN(Time Delay Network)이다. TDNN은 음성이 특징들을 잘 분류할 수 있게끔 MLP의 입력단에 시간적인 특징도 포함되도록 한 알고리즘으로서 음소, 단어에 높은 인식률을 보인다. 또 다른 알고리즘으로서 LVQ(Learning Vector Quantization)가 있다. LVQ는 Kohonen이 제안하였는데 인간 뇌의 특성을 고전적인 VQ방식에 적용한 것으로서 음성 인식 시스템에 이용할 경우 높은 인식률을 보인다.

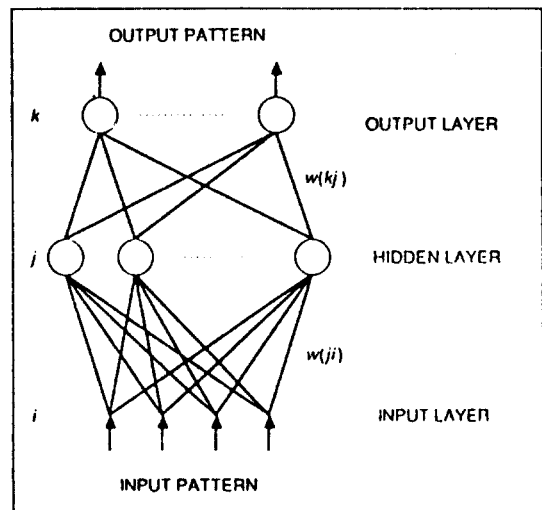


그림 1. 신경회로망의 구성

Fig. 1. A schematic depiction of a semilinear feedforward connectionist net.

MLP의 학습은 회로망의 현재출력과 기대출력의 차이를 소멸시키는 방향으로 결합강도를 조절하는 것으로 본 실험에서는 Rumelhart등에 의해 제안된 오차 역접파법(Back Propagation)을 사용하였다.

순방향 다층 신경회로망은 다수의 시그모이드 유닛으로 이루어진다. N층의 회로망에서, 제n층 i번째 유닛의출력을 $S_i^{(n)}$ 으로 하면, 전향 관계는 다음식에 의해 나타내진다.

$$net_1^{(n)} = \sum_{i \in I_1} W_{ij}^{(n)} S_j^{(n-1)} + W_{i0}^{(n)} \quad (3-3)$$

$$S_i^{(n)} = \frac{1}{1 + \exp(-net_1^{(n)})} \quad (3-4)$$

로 된다. 여기서 $n=1,2,3,\dots,N$ 이고, $W_{i0}^{(n)}$ 은 바이어스(bias) 값, $W_{ij}^{(n)}$ ($0 < j$)는 웨이트로 불린다. 주어진 입력패턴 (i_{pi})에 대해, $S_i(0)=i_{pi}$ 로 하고, 식(3-3), 식(3-4)에 따라서 각 층의 $net_1^{(n)}$, $S_i^{(n)}$ 출력층에서 회로망의 출력패턴 $S_j^{(n)}$ 을 구한다. 여기서 교사신호 (T_{pj})와의 오차 $e_j = [S_j^{(n)} - (T_{pj})]$ 의 자승합수

$$E_p = 1/2 \sum_j (e_j)^2 = 1/2 \sum_j (S_j^{(n)} - T_{pj})^2 \quad (3-5)$$

을 회로망의 평가함수로 한다. 평가함수의 제n층 i번째 유닛의 $net_1^{(n)}$ 에 관한 편미분을

$$\frac{\partial E_p}{\partial net_1^{(n)}} = \delta_i^{(n)} \quad (3-6)$$

로 하고, 다음 식을 사용하여 제n층의 $\delta_i^{(n)}$ 을 역전파하여, 제n-1층 j번째 유닛의 $\delta_j^{(n-1)}$ 을 구한다.

$$\delta_i^{(n-1)} = S_j^{(n-1)}(1-S_j^{(n-1)}) \sum_{n=N, N-1, \dots, 2} \delta_i^{(n)} W_{ij}^{(n)} \quad (3-7)$$

여기서 $\delta_i^{(n)} = S_j^{(n)}(1-S_j^{(n)})e_j$ 이다.

웨이트 W_{ij} 에 관한 편미분은 다음 식에 의해 계산한다.

$$\frac{\partial E_p}{\partial W_{ij}^{(n)}} = \frac{\partial E_p}{\partial net_1^{(n)}} \frac{\partial net_1^{(n)}}{\partial W_{ij}^{(n)}} = \delta_i^{(n)} S_j^{(n-1)} \quad (3-8)$$

바이어스값 $W_{i0}^{(n)}$ 에 관한⁶ 편미분은, 뒷 식에서 $S_j^{(n-1)} = 1$ 로서 구한다.

Rumelhart는 위에 나타낸 오차역전파 알고리즘을

사용해서 웨이트와 바이어스값에 관한 遍微分을 계산하고, 평가함수 E_p 를 최소로 하는 각 유닛의 웨이트와 바이어스값을 새로운 식

$$\Delta W_{ij}(n) = -\eta \frac{\partial E_p}{\partial W_{ij}} + \alpha \Delta W_{ij}(n-1) \quad (3-9)$$

에 의해, 반복적으로 구한 방정식을 제안했다. 식(3-7)에서 η 와 α 는 0에서 1사이의 값이 되고, 각각 학습률(learning rate), 관성률(momentum factor)로 불리우고 있다. 관성률 α 는 편미분의 진동을 억누르며 진동하지 않는 부분을 확대하는 효과가 있다. 여기서 $\eta < \lambda_{max}$ (λ_{max} 는 웨이트와 바이어스값에 관한 2차 편미분 행렬)일때 알고리즘은 수렴한다.

또한, 통상의 웨이트와 바이어스값의 초기치는 $(-\delta, +\delta)$ 범위의 난수를 준다.

식(3-7)의 $\frac{\partial E_p}{\partial W}$ 의 대응으로,

$$\frac{\partial E}{\partial W} = \sum_p \frac{\partial E_p}{\partial W} \quad (3-10)$$

을 사용할 수도 있다. 여기서, $E = \sum E_p$ 는 모든 패턴의 총오차자승합이된다.

2-3. NN-HMM에 의한 단어인식 시스템

HMM에서 개인차에 의한 변동을 충분히 흡수하기 위해서는 다량의 데이터를 학습 데이터로 하여 개인차에 의한 오인식을 다소 보완할 수 있는데 본 연구에서는 소량의 학습 데이터로 개인차에 의한 애매한 스펙트럼 특성을 흡수할 수 있도록 HMM과 신경회로망을 각각 독립적인 사건으로 가정하여 두 시스템의 출력을 하나의 사건으로 수정하여 인식하는 방법으로 각각의 시스템의 결점을 보완하여 인식률을 개선한 NN-HMM 방식을 제시하고자 한다.

본 연구 방법에는 HMM의 인식결과와 인덱스 출력확률들을 입력으로 사용한 신경회로망이 상호 보완적인 연결로 서로의 오인식을 보정해 줌으로써 효율적인 인식기를 구현하는데 있다. HMM에서 Viterbi 알고리즘으로 인식시 가장 큰 최종 확률값의 음성모델로 인식을 행하므로 만약 어느 음성모델의 확률이 개인차에 의해 근소한 차로 오인식의 결과를 초래한다면 이를 보정하는 기능이 없다. 그리고 본 연구에서 이용되는 신경회로망은 인식시 음성의 발생시간에 대한 변동을 흡수하지 못하므로 저조한 인식 결과를 가져다온다. 이런 두가지의 독립적인 인식시스템

을 하나의 시스템으로 구성하여 HMM과 신경회로망의 오인식의 주요한 원인인 개인차 또는 발성 시간의 변동량을 상호 보정하여 효과적인 음성인식을 수행하는 것이다. 이 방법은 두개의 개별적인 시스템의 출력을 독립적으로 보아 인식하는 음성모델은 두 시스템 출력의 확률곱으로 인식하는 방법이다.

인식 시스템 구성은, 음성의 특징을 추출한후에 벡터양자화에 의해 코드북을 만들고 이에 따른 인덱스로 양자화 시킨다. 이 대표 인덱스는 HMM과 신경회로망의 입력으로 사용하여 학습과 인식을 수행한다. 본 연구에서의 인식 시스템 구성은 그림 2와 같이 구성된다.

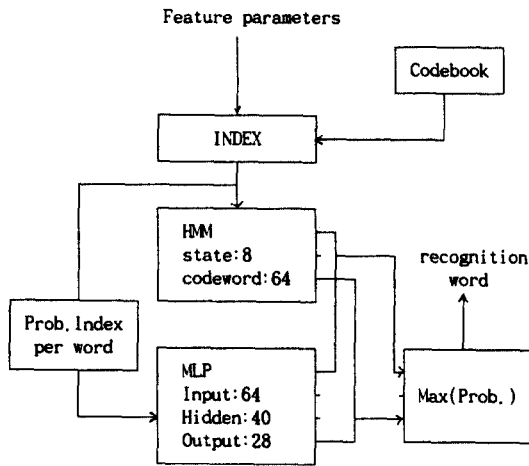


그림 2. 음성 인식 시스템 구성도
Fig. 2. Block diagram of an isolated word recognizer

벡터 양자화를 행한후 각 음성모델은 프레임마다 코드북과의 거리가 최소인 대표 인덱스를 구하여 그 인덱스를 HMM과 신경회로망의 입력으로 한다. 학습은 HMM과 신경회로망이 각각 서로 독립적으로 학습을 행한 후 인식과정은 각 음성모델의 신경회로망 출력값과 HMM의 Viterbi알고리즘으로의 출력 확률을 갖는 음성모델을 인식하도록 구성하였다.

HMM은 실험을 통하여 hidden 상태가 8이고 코드북 크기는 64일때의 인식률이 가장 좋은 결과를 나타내었다. 따라서 신경회로망에는 같은 조건으로 모델마다의 출력 인덱스 확률을 입력 파라미터로 사용하였다. 그리고 신경회로망의 입력층은 64, 중간층은 40, 출력층은 28의 3층 신경회로망을 이용하고, 역전

파법에 의해 학습시킨다. 이때 역전파의 학습파라미터는 학습률은 0.3, 관성률은 0.9로 하였고, 오차한계는 0.001로 하였다.

III. 실험 결과 및 고찰

본 연구에 쓰인 음성 데이터는 방음장치가 되지않은 연구실에서 20대 남,여 화자에 의해 자연스럽게 발음된 DDD 지역명을 마이크를 사용하여 3.5KHz의 LPE(Lowpass Filter)를 통과 시킨 후 10KHz의 샘플링 주파수로 A/D변환기를 통해 12비트로 양자화하였다. 이 음성 데이터는 끝점 검출을 행한 음성신호 s(n)을 프리엠퍼시스처리 후 윈도우는 Hamming 창을 사용하였다. 음성 분석은 20.0ms(200샘플)를 1프레임으로 이동 구간은 10.0ms(100샘플)로 50% 중첩하였고, 벡터양자화를 위한 특징벡터로는 14차 선형예측계수 벡터로 하였다. 이 중에서 벡터양자화와 코드북을 생성하기 위한 음성 신호로는 5회 발음한 것 중에서 1회 발음한 28개의 지역명에서 추출한 1446개의 벡터로 하였다.

본 연구에서 아용된 데이터는 특별시와 직할시 그리고 경기도 지역의 DDD 지역명 28로써 이 중 남,여 화자가 3번 발음한 총 252개의 데이터로 인식 실험을 하였다. 학습에 참여한 데이터는 인식 실험에는 사용을 하지 않았다. 본 연구에서는 제안하는 방법 이외에도 비교를 위하여 동일한 데이터와 조건으로 전통적인 이산분포 HMM에 의한 방법과 신경회로망에 의한 방법도 실험하여 비교 평가하였다.

3-1. 실험인식 결과

3-1-1. HMM에 의한 인식 결과

학습에 참여한 3명의 화자가 5회 발성한 데이터중

표 1. 코드워드에 따른 HMM의 인식률. (상태=8)
Table. 1. Recognition rate of HMM for codeword. (state=8)

화자	코드워드 수		
	32	64	128
MA	82	92	61
WB	79	88	48
MC	92	93	73
AVG.	85	91	61

MA : 남성 화자1 WB : 여성 화자 MC : 남성 화자2

표 2. 상태 수에 따른 HMM의 인식률. (코드워드=4)
Table. 2. Recognition rate of HMM for states.
(codeword=64)

상태수 화자	4	5	6	7	8	9
MA	76	84	89	92	92	92
WB	80	83	88	88	88	88
MC	83	86	92	92	93	93
AVG.	80	84	90	91	91	91

3회 발성한 데이터를 학습을 하고 나머지 2회 발성한 음성 데이터를 코드북의 크기에 따른 HMM의 인식률은 표1과 같다. 단 그때의 HMM의 hidden 상태는 8로 하였을 경우이다. 또 표2는 코드북 사이즈를 64로 하여 HMM의 hidden 상태를 가변하여 실험한 결과이다.

위의 실험결과에서 HMM의 상태는 8, 코드북의 크기는 64 인 경우에 가장 적합한 방법으로 생각되며 이에 따라 신경회로망의 실험에는 64 코드워드를 이용하여 실험하였다.

3-1-2. 신경회로망에 의한 학습 결과

본 연구에서는 신경회로망의 입력층은 64, 중간층은 40, 출력층은 28의 3층 신경회로망을 이용하였고 역전파법에 의해 학습시켰다. 이때 역전파의 학습파라미터는 학습률은 0.3, 관성률은 0.9로 하였고, 오차한계는 0.001로 하였다. 이에 따르는 학습 오차는 그림3에 나타내었다.

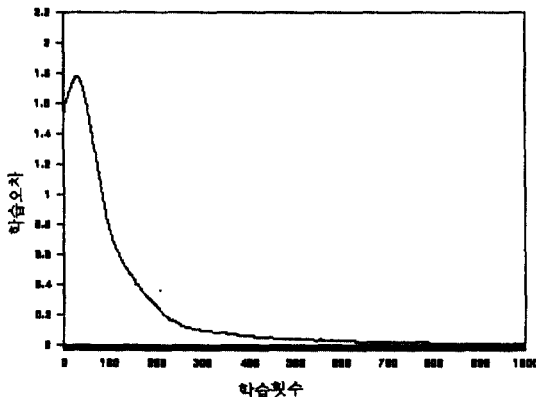


그림 3. 신경회로망의 학습 오차
Fig. 3. System Error of the Neural Network.

3-1-3. NN-HMM에 의한 인식 결과

표3.은 신경회로망과 HMM이 결합하여 하나의 시스템으로 구성하였을 경우의 인식률이다. 코드워드는 64로 하여 실험한 결과이다.

표 3. 인식률의 비교 (codeword=64)
Table. 3. Comparison of recognition rates.
(codeword=64)

인식시스템 화자	HMM	NN	NN-HMM
MA	92	90	95
WB	88	87	95
MC	93	90	96
AVG.	91	89	95

위 결과를 보면 HMM과 신경회로망에 의한 인식율보다 NN-HMM에 의한 인식율 향상이 남성화자보다 1명의 여성 화자의 인식율 향상이 높은 것을 알수 있으며, 오인식 보정이 2명의 남성화자 보다 더욱 잘 이루어 진다는 것을 알수있다. 이는 적은 수의 여성 화자로 잠성화자와 거의 대등한 인식률을 나타내고 있는데 이는 성별차에 의한 개인차를 흡수함으로 높은 인식결과를 얻을수 있음을 나타내고 있다고 할 수 있다.

IV. 결 론

본 연구에서는 NN-HMM 학습방법을 이용하여 보다 효율적인 음성인식 시스템을 구현하는데 있다. 이 제안된 음성인식 시스템으로 인식실험을 수행한 결과 이산분포 HMM보다 약 4[%]의 인식률을 개선할 수 있었고 신경 회로망을 이용한 방법보다 약 6[%]의 인식률을 개선할 수 있었다. 이런 실험 결과는 서로 다른 특성의 음성인식 시스템이 오인식을 서로 보정해 줌으로써 높은 인식률을 나타낼수 있었다. 이러한 음성인식 시스템은 소량의 학습데이터로 효율적인 결과를 가져와 대,중규모의 단어 인식 시스템에 적합한 방법으로 사료되고 다음과 같은 결론을 얻을 수 있었다.

첫째, NN-HMM은 벡터 양자화의 대표인덱스를 입력으로 함으로서 시스템의 구현과 전처리를 간단히 구성할 수 있었다.

물째, 전통적인 HMM은 높은 인식률을 위하여 많은 학습 모델이 반드시 필요로 하는데 NN-HMM은 적은 학습 데이터로서 높은 인식결과를 얻을 수 있었고 또한 인식시 필요한 시간도 기존의 HMM의 인식 시간과 차이가 없었다.

셋째, NN-HMM은 HMM과 신경회로망이 상호 보정작용에 의하여 남,여간의 성별차에 의한 오인식률이 보다 감소함으로써 높은 인식률을 얻었다.

참 고 문 헌

1. A. Waibel et al. "Phoneme recognition using time-delay neural networks". IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, Mar. 1989, pp. 328-339
2. K. C. Pan, F. K. Soong, L. R. Rabiner, "A vector-quantization-based preprocessor for speaker-independent isolated word recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, NO.3, pp. 546-553, 1985
3. J. G. Wilpon, L. R. Rabiner, "A modified K-means clustering algorithm for use in isolated word recognition," IEEE Trans. Acoust. Speech, Signal Processing., Vol. ASSP-33, No. 3, pp. 587-595, 1985
4. Y. Linde, A. Buzo, R. M. Gary, "An algorithm for vector quantization design," IEEE Trans. Communication, Vol. COM-28, pp. 84-108, 1980
5. L. R. Rabiner, "On the application of vector quantization and hidden Markov models to speaker independent isolated word hidden Markov models to speaker independent isolated word recognition," BSTJ, Vol. 62, No. 4, Apr. 1983
6. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of IEEE, Vol. 77 No. 2, Feb. 1989
7. R. P. Lippman, "An introduction to computing with neural nets," IEEE ASSP Magazine, Apr. 1987
8. B. R. Kammerer, W. A. Kupper, "Experiment for isolated-word recognition with single and two layer perceptrons," Neural Networks, Vol. 3, pp. 693-706, 1990



金 鉉 秀 (Yeun Soo Kim) 正會員
 1963年 10月 5日生
 1989年 2月 : 明知大學校 工科大學
 電子工學 卒 (工學士)
 1992年 8月 : 明知大學校 大學院 電
 子工學科 卒 (工學碩士)



金 昌 錫 (Chang Seok Kim) 正會員
 1938年 9月 18日生
 1968년 2월 : 명지대학교 전기공학과 졸업 (공학사)
 1972년 2월 : 명지대학교 대학원 전기공학과 졸업 (공학석사)
 1985년 2월 : 명지대학교 대학원 전자공학과 졸업 (공학박사)

1985년 3월 ~ 현재 : 명지대학교 공과대학 전자공학과 교수