

신경망을 이용한 모음의 학습 및 인식 방법

(A Method of Learning and Recognition of Vowels by Using Neural Network)

沈在亨*, 李鍾赫**, 尹台焄*** 金在昌*** 李梁成***

(Jae Hyoung Shim, Jong Hyeok Lee, Tae Hoon Yoon, Jae Chang Kim, and Yang Sung Lee)

要 約

본 연구에서는, Ohotomo 등이 모음의 학습과 인식을 위해 구성한 BP 구조 신경망의 학습을 위해 사용하였던 입력 패턴의 방법을 보완하여, 포먼트 주파수의 대역폭을 고려한 측면값을 학습용 입력패턴에 두어 수렴 속도와 인식율을 높이고자 한다. 본 연구에서 제안한 방법이 오인식율에서는 30% 정도의 감소와 수렴 속도면에서는 7%의 증가를 컴퓨터 시뮬레이션을 통하여 알 수 있었다.

Abstract

In this work Ohotomo et al., neural network model for learning and recognizing vowels is modified in order to reduce the time for learning and the possibility of incorrect recognition. In this modification, the finite bandwidth of formant frequencies of vowels are taken into considerations in coding input patterns. Computer simulations show that the modification reduces not only the possibility of incorrect recognition by about 30% but also the time for learning by about 7%.

I. 서 론

인간과 기계 사이의 효율적인 정보 전달을 위해서 시작된 음성인식에 관한 연구는 컴퓨터 및 통신 정보 기술의 발전에 힘입어 많은 발전을 이루었다.

1980년대에 접어들면서 벡터 양자화(vector quantization), HMM(hidden markov model), 그리고 DTW(dynamic time warping) 등의 방식을 도입하여 격리 음성과 화자 종속 등의 경우를 대상으로 연구한 결과 실용화 단계에 있으나, 연속 음성과 화자 독립의 경우는 조음결합과 음소, 음절 및 단어간의 경계를 분리하여야 하는 문제들로 인하여 실용화하는데 많은 어려움이 따른다.^(1,2)

최근 신경망에 관한 연구의 급진적으로 음성 인식에 유용한 자율조직 특성(self organization), 훈련 특성(training), 그리고 적응특성(adaptation) 등의 여러 특성이 밝혀지고 있고, 이를 연속 음성, 화자 독립의 경우에 적용하려는 연구가 활발히 진행되고 있다. 끝

*準會員, 曉星電子研究所
(Electronic Research Lab., Hyosung Computer)

**正會員, 慶星大學校 컴퓨터工學科
(Dept. of Computer Eng., Kyungsung Univ.)

***正會員, 釜山大學校 電子工學科
(Dept. of Elec. Eng., Pusan Nat'l Univ.)

接受日字: 1990年 2月 9日

드(gold)는 홉필드(hopfield) 모델을 사용하여 모음 인식을 하였으며,^[4] 리프만(lippmann)은 BP(back propagation) 학습 알고리즘을 사용하여 숫자 음성을 인식하였다.^[4] 코호넨(kohonen) 등은 음성의 포만트(formant) 주파수를 입력으로 하는 자율조직 신경망을 구성하여 음성 인식을 하였으며, 이를 이용하여 음성 타자기(phonetic typewriter)를 구성하였다.^[2] Ohotomo 등은 5개의 모음에서 얻은 제1포만트 주파수와 제2포만트 주파수 정보를 이용하여 음운 지각 분포도를 만들었다. 이를 근거로 BP 구조를 갖는 신경망의 학습에 교사신호로 사용할 100개의 입·출력 패턴을 만들었고, 이 중 80개를 선택하여 학습시킨 후, 100개의 입력 패턴을 제시하였을 때 100개 모두 인식할 수 있다고 보고하였다.^[5] 국내에서는 hopfield network와 BP 신경망을 이용한 모음의 인식을 행하고 있으며 대부분 기존의 방법을 한국어의 특징에 맞도록 적용시키고 있다.^[6-8]

한편 BP 학습 알고리즘의 경우 수렴 속도, 학습 성취도는 신경망의 초기 가중치(weight)의 설정 상태와 학습 대상의 선정에 크게 좌우되는 것으로 알려져 있다.^[9] 학습 성취도를 높이는 방법으로 잡음에 의해 변형된 입력 패턴을 인가하는 방법, 오인식된 패턴만을 모아서 선택적으로 재학습시키는 방법, 학습과정시 구조적인 정보를 이용하는 법이^[10] 있으며, 수렴 속도를 높이는 방법으로 모멘텀항(momentum term)을 도입하여 학습시키는 방법 등이^[9] 제안되었지만 이와 같은 문제를 해결하기 위한 많은 연구는 계속되고 있다.

본 연구에서는 모음의 음운 지각 분포도를 신경망에 학습시키는데 있어 수렴 속도와 학습 성취도를 높이는 방법으로 기본 모음의 제1, 제2포만트 주파수의 대역폭을 고려하여 설정한 측면값(side value)을 학습용 입력 패턴에 두는 방법이 사용될 수 있음을 컴퓨터 시뮬레이션을 통하여 보이고자 한다.

II. 신경망의 구성과 학습 알고리즘

사람의 음성 인식과정은 음성 파형에서의 음소 검출로 시작하여, 메시지에 대한 이해에 이르는 여러 과정으로 이루어져 있다. 그리고 사람이 음성을 인식할 때는 주어진 음성의 모든 요소를 듣는 것보다 많은 경험에 의해서 익숙해진 전후 문맥에 부분적인 음성 신호를 계속적으로 관련시켜 전체 의미를 파악하게 된다. 음성 인식 시스템이 고도의 인식 기능을 갖추기 위해서는, 부가적으로 음성 신호의 어의(semantics)를 이해할 수 있어야 하는 것으로 알려

져 있다.^[2] 그러나 기본 모음을 인식 대상으로 할 경우 제1, 제2포만트 주파수 정보만으로도 가능한 것으로 알려져 있다. 본 장에서는 모음의 음운 지각 분포도를 BP 학습 알고리즘을 사용하여 신경망에 학습시키는데 ohotomo 등이 사용한 방법^[5]과 본 연구에서 제안한 방법을 수렴 속도와 학습 성취도의 관점에서 서로 비교하기 위한 모음의 음운 지각 분포도, 학습과정 중 사용될 입·출력 패턴의 구성, 신경망의 구성과 뉴런의 특성 및 BP 학습 알고리즘에 대하여 기술하였다.

1. 모음의 음운 지각 분포도

Ohotomo 등은 5가지 모음(a, i, u, e, o)을 BP 신경망에 학습시키기 위해서, 모음의 제1, 제2포만트 주파수를 변화시키면서 얻은 합성 음성의 음운 지각 분포도를 그림 1과 같이 포만트 주파수 평면상에 나타내었다. 가로축은 제1포만트 주파수를 세로축은 제2포만트 주파수를 나타낸다. 그리고 그림 1에 나타난 모음의 포만트 주파수 범위는 표 1과 같다.

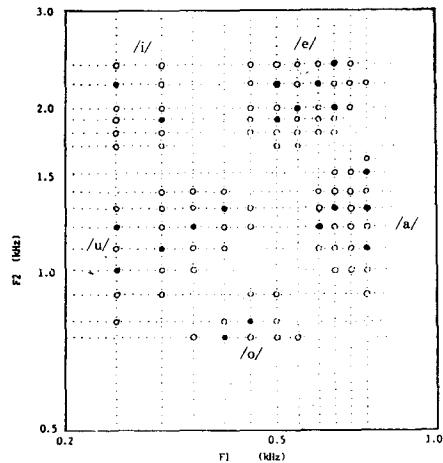


그림 1. 기본 모음의 음운 지각 분포도
Fig. 1. Perceptual phoneme distribution of fundamental vowels.

표 1. 모음의 포만트 주파수 범위
Table 1. The range of formant frequency of vowels

모음 종류	갯수	제 1 포만트 주파수 [Hz]	제 2 포만트 주파수 [Hz]
/a/	24	400~800	900~1600
/e/	31	450~750	1700~2400
/i/	12	250~300	1700~2400
/o/	10	350~550	750~900
/u/	23	250~450	800~1400

Ohotomo 등은 BP신경망을 이용하여 모음의 음운 지각 분포도를 학습시킬 때, 그 일부만 사용하여 학습시킨 후 학습결과를 평가하였으며, 학습과정 중 제시하지 않은 입력 패턴들도 신경망의 학습 특성으로 인하여 인식할 수 있음을 보였다. Ohotomo 등은 이 실험을 위해서 총 100개의 모음 음운 지각 분포도상의 데이터 중 임의로 80개를 선택하여 그림 1에 표시된 바와 같이 흰점으로 표시하였고, 나머지 20개는 검은점으로 표시하였다. 여기서 흰점 80개만 신경망을 학습하는데 사용하였으며, 학습 결과의 평가는 학습 과정중에 제시하지 않았던 20개의 패턴을 포함한 총 100개의 패턴을 제시하여 학습 성취도를 평가하였다. 본 연구에서는 그림 1의 음운 지각 분포도상의 100개의 패턴 가운데 80개, 60개, 40개씩 임의로 선택하여 학습용 패턴수를 변화시키며 학습시킨 후 학습결과를 평가하였다.

2. 입·출력 패턴의 구성

BP 구조의 신경망을 학습시키기 위해서는 외부교사(supervisor)가 필요하다. 외부교사는 입력패턴과 이에 대응하는 바람직한 출력 패턴을 구성하며, 이를 신경망에 반복 제시하여 신경망으로 하여금 지식을 학습하도록 한다. Ohotomo 등은 입력 패턴을 구성하기 위해서 음성 신호의 포먼트 주파수를 표 2와 같이 분할하여 각 구간을 입력 패턴의 각 비트에 대응시켰다. 입력 패턴의 길이를 총 26개의 비트로 구성하였으며 포먼트 주파수가 속하는 구간에 대응하는 입력 패턴의 비트를 "1.0"으로 두었고, 그 이외 비트에 대해서는 "0.0"으로 하여 각 모음의 입력 패턴을 구성하였다.

표 2. 포먼트 주파수의 분할간격
Table 2. The division interval of formant frequency of vowels.

포먼트 주파수 범위	분할 간격
250~800Hz	50Hz 간격
800~2,000Hz	100Hz 간격
2,000~2,400Hz	200Hz 간격

그림 2는 모음 /e/에 대한 주파수 응답특성을 나타낸 것이다.¹¹⁾ 그림 2에서 제1, 제 2 포먼트 주파수의 대역폭은 각각 50Hz, 100Hz 정도이다. 포먼트 주파수의 인접 주파수에서 진폭은 공진회로의 대역폭이 포먼트 주파수에서 대역폭과 같으며 중심 주파수가

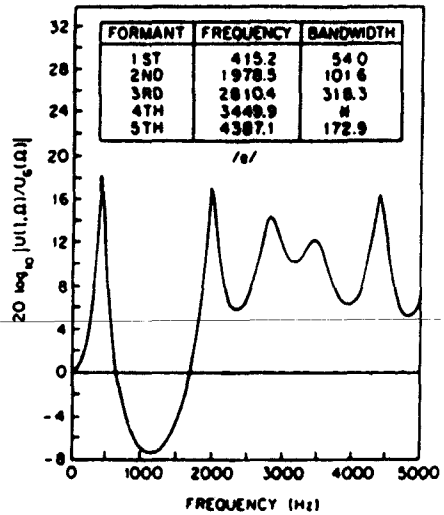


그림 2. 모음 /e/에 대한 주파수 응답 특성
Fig. 2. Frequency response for a vowel /e/.

포먼트 주파수와 같은 RLC 공진 특성에서 구할 수 있다. 중심주파수가 속하는 구간의 좌우 입력 패턴의 비트, 즉 주파수 영역상의 제 1 포먼트 주파수에서 50Hz, 제 2 포먼트 주파수에서 100Hz~200Hz정도 떨어진 주파수에서의 진폭을 RLC 공진회로의 전달함수를 이용하여 계산하면 약 "0.42~0.47" 정도의 값을 가짐을 알 수 있다. Ohotomo 등은 이 값을 무시한 채 "0.0"으로 취급하여 학습용 입력 패턴을 구성하였다. 본 연구에서는 입력 패턴의 구성시 이 점을 고려하기 위하여, 포먼트 주파수가 속한 구간에 대응하는 입력 패턴의 비트에 "1.0"의 값을 할당하였으며, 포먼트 주파수가 속하는 구간 좌우 입력 패턴의 비트에는 "0.0~0.47"사이의 적절한 값을 할당하였다. 이렇게 함으로써 입력패턴의 구성시 포먼트 주파수의 대역폭을 고려하도록 하였으며, 이 방법이 수렴 속도와, 학습 성취도에 어떤 영향을 미치는가 알아 보았다. 출력 패턴의 길이는 인식하고자 하는 모음의 유형의 수 만큼인 5비트를 두었고, /a/, /e/, /i/, /o/, /u/의 차례로 각 모음을 출력 패턴의 각 비트에 대응시켜 구성하였다. 출력 패턴은 5가지 모음에 해당하는 임의의 입력 패턴이 신경망에 입력되었을 때, 이 입력 패턴에 대응하는 모음만 선택되도록 출력 패턴을 구성하였다. 즉 선택된 모음에 해당하는 비트만 "1.0"로, 그 이외의 비트는 "0.0"으로 코우딩하였다. 예를 들어 /o/에 해당하는 음성신호(F1=500Hz, F2=800Hz)를 코우딩할 경우 ohotomo 등은

입력 패턴을 1000001000001000000000000001로 구성 하였으나,^[6] 본 연구에서는 대역폭을 고려하여 입력 패턴을 10000*1*000*1*000000000000001로 구성 하였다. 여기서 *표시는 축면값을 나타내며 “0.0~0.47”사이의 값으로 실험적으로 구한 최적의 값을 나타낸다. 그리고 바라는 출력 패턴은 두 경우 모두 1000101로 구성하였다.

3. 신경망의 구성 및 뉴런의 특성

그림 3은 그림 1과 같은 모음 음운 지각 분포도를 학습시키기 위하여 본 연구에서 사용한 BP 구조를 이룬 신경망으로 출력층과 입력층, 그리고 하나의 히든층을 갖는 3층 구조를 이루고 있다.

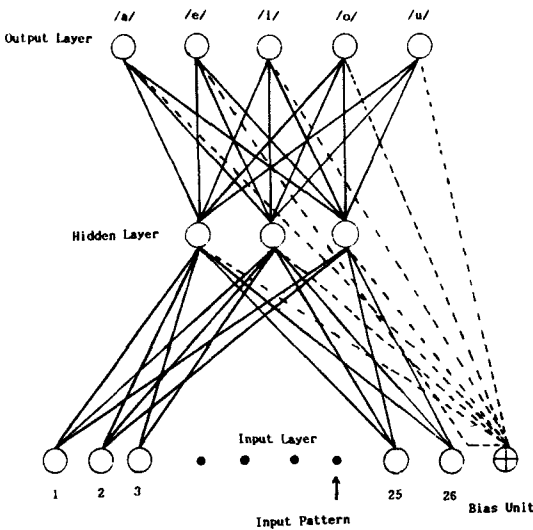


그림 3. 음운 분포를 학습시키기 위한 신경망 모델
Fig. 3. A neural net model for learning perceptual phoneme distribution.

히든층의 뉴런수는 이론적으로 몇 개까지라도 들 수 있지만, 일반적으로 원하는 출력이 다섯가지일경우 2~5개 정도 둔다. 입·출력층에 위치한 뉴런은 외부와 직접적으로 연결되어 있으며, 히든층에 위치한 뉴런은 외부와 직접적으로 연결되어 있지 않다. 그리고 같은 층에 위치한 뉴런 간의 축방향의 연결은 고려하고 있지 않다. 그림 4는 BP 네트워크에서 사용되는 뉴런의 구조를 나타낸 것이다.

신경망은 기본적으로 노드(node)와 이들간을 서로 연결하는 연결 가중치로 구성된다. 노드는 뉴런 또는 PE(processing element)라고 하는 개별처리 소

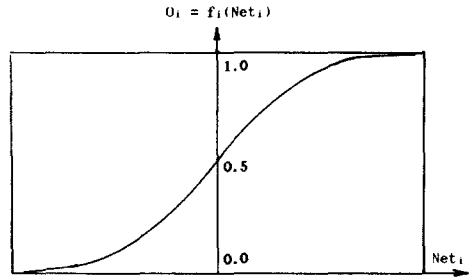
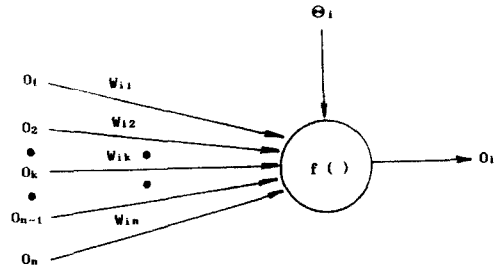


그림 4. BP 네트워크에서 사용되는 뉴런의 구조
Fig. 4. Structure of neuron used in BP network.

자들 가지며, 이들은 서로의 연결을 통하여 정보를 주고 받는 병렬 분산 처리하는 구조를 갖는다. 뉴런 U_i 의 총 입력량 Net_i 는 그림 4와 같이 다른 뉴런 U_k 의 출력 O_k 와 뉴런 U_i 와 U_k 간의 연결세기 W_{ik} 로 부터

$$Net_i = \sum_k W_{ik} O_k + \theta_i \quad (1)$$

로 표시된다.

여기서 θ_i 는 뉴런 U_i 의 bias를 나타낸다.

뉴런망의 전달함수에 식(1)에서 구한 총 입력량을 인가하면 뉴런 U_i 의 실제 출력 O_i 는 다음과 같다.

$$O_i = f_i(Net_i) = \frac{1}{1 + \exp(-Net_i)} \quad (2)$$

여기서 뉴런망의 전달함수는 식(2)와 같이 미분가능하며 단조증가 함수인 sigmoid함수를 사용하였다.

4. BP신경망의 학습 알고리즘

BP신경망의 학습 알고리즘은 Rumelhart^[9] 등이 “Generalized delta rule”을 일반화한 것으로, 학습을 위해서 외부교사가 존재하여 입력 패턴 x 에 대해 원하는 출력 패턴 t 를 출력하도록 연결 가중치를 조정하여 실제 출력 o 와 원하는 출력 t 와의 오차 e 를 최소로 하는 것이다. BP신경망에서 뉴런간의 연결 세

기를 w 라 하고, 학습시킬 대상 패턴을 $p=1, 2, 3, \dots, n$ 라고 하자. 그리고 입력 패턴 p 에 해당하는 원하는 출력 패턴과 실제의 출력 패턴을 각각 t^p, o^p 라고 하자. 그리고 각 패턴 p 에 대응하는 에러를

$$e^p = (t^p - o^p) \tag{3}$$

로 정의할 때, 패턴 p 에 대한 cost 함수는

$$E_p(w) = ||e^p||^2 = ||t^p - o(w)||^2 \tag{4}$$

로 표시되며, 전체 패턴에 대응하는 cost 함수는

$$E(w) = \sum_p E_p(w) \tag{5}$$

로 주어진다.

신경망을 학습시킨다는 것은 $E(w)$ 를 최소로 하는 w 를 찾는 문제로 귀착되는데, Rumelhart 등은 급강하법(steepest descent method)을 사용하여 다층 인식자(multilayer perceptron)의 $E(w)$ 를 최소로 하는 w 를 반복적으로 찾는 방법을 다음과 같이 유도하였다.

$$W_p(n+1) = W_p(n) + \Delta W_p(n+1) \\ \Delta W_{ij}(n+1) = \eta \delta_i O_j + \alpha \Delta W_{ij}(n) \tag{6}$$

$$\theta_p(n+1) = \theta_p(n) + \Delta \theta_p(n+1) \\ \Delta \theta_i(n+1) = \eta \delta_i + \alpha \Delta \theta_i(n) \tag{7}$$

$$\delta_{pi} = f'_i(\text{Net}_i) (t_i - o_i) \\ f'_i(\text{Net}_i) = \sum_k W_{ki} \delta_k \tag{8}$$

여기서 δ_i 는 후방향(backward)으로 전달되는 에러를 나타낸다. 바이어스항 θ 는 입력값이 항상 "1.0"인 뉴런과 다른 층의 뉴런간을 서로 연결하는 연결세기로 취급하여 θ 도 w 와 마찬가지로 학습시킬 수 있다. η 는 학습율(learning rate)을 나타내는 상수이며, η 는 $0 < \eta < 1$ 의 범위를 만족하는 값으로, 시뮬레이션에서 적절한 값으로 두며, α 는 모멘텀항 즉 현재의 가중치 변화량을 결정하는데 과거 가중치 변화량을 반영하는 상수이다.

III. 시뮬레이션 결과 및 검토

1. 시뮬레이션

본 시뮬레이션을 위해서 그림 3 과 같은 BP신경망을 구성하였다. 입력층의 뉴런수는 27개이고, 이 중 26개를 입력 패턴의 각 비트에 대응시켰으며, 입력 패턴의 각 비트값이 입력층의 입력값이 되도록 하였다. 나머지 1개의 뉴런은 식(1)의 바이어스항을 처리하기 위해서 도입하였다. 히든층의 뉴런 개수를 2~

5개까지 변화시키면서 시뮬레이션하고, 이 결과에서 얻어지는 학습시간과 하드웨어 구성의 복잡도를 고려하여 히든층의 뉴런수를 결정하였다. 인식하고자 하는 모음의 갯수가 5개이므로 출력층의 뉴런 수는 5개로 하였으며 출력에서 0.1이하는 "0.0"으로, 0.9 이상은 "1.0"로 하여 출력패턴을 구성하고, 원하는 출력 패턴과 비교하여 인식여부를 결정하였다. 신경망의 초기 연결 가중치 W_{ij} 는 $-0.3 < W_{ij} < 0.3$ 을 만족하는 난수로 설정하였다. 그리고 학습 성취도는 신경망의 초기 가중치의 설정상태와 학습대상의 설정에 따라 크게 좌우되므로 학습과정의 처음 시작때 신경망의 초기 가중치를 바꾸어가며 신경망을 초기화 하였으며, 학습대상 모음의 선정은 그림1의 모음은 지각 분포도상의 100개의 샘플중 임의로 80개, 60개, 40개씩 선택하여 학습시켰다. 실험 결과는 초기 조건들을 바꾸어가며 얻은 데이터를 근거로 평가하였으며 알고리즘은 IBM-286 호환기종에서 MSC를 사용하여 구현하였다.

2. 히든층의 뉴런 수 결정

그림 5는 입력 패턴을 ohotomo 등의 코오딩 방법^[5]과 같이 구성하고 히든층의 뉴런수를 2~5개까지 변화하면서, 반복 학습에 따른 t_{ss} (total sum of square of error)의 변화를 나타낸 것이다. 그림5의 가로축은 학습용 패턴의 반복제시 횟수를 나타내며, 세로축은 $\text{Log}_{10}|t_{ss}|$ 로 나타내었다. 출력 뉴런이 N 개 일때 히든층의 뉴런수가 $\lceil \log_2 N \rceil$ 개 이상 되었을 때에 수렴 속도가 빠름을 알 수 있다. t_{ss} 값이 0.05 이하로

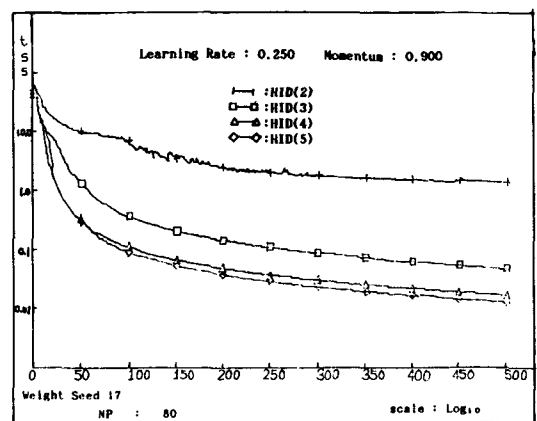


그림 5. 히든층의 뉴런수를 2~5개까지 변화시킬 때의 수렴특성
Fig. 5. Convergence of t_{ss} for learning iteration, as the number of hidden layer is varied.

될 경우를 수렴의 기준으로 했을 때, 뉴런의 수를 2개로 하였을 경우 잘 수렴하지 않았으며, 뉴런의 수가 3개 이상이 될 때 수렴하였다. 뉴런의 갯수를 4개, 5개로 증가시킬 때 시뮬레이션에서 수렴에 필요한 반복 횟수가 많이 감소하므로 전체 학습 시간은 감소하였다. 그러나 실제 시스템을 하드웨어로 구성할 경우 히든층의 뉴런 수가 증가할수록 시스템 구성이 더욱 복잡해지므로 본 연구에서는 히든층의 뉴런수를 3개로 하였다.

3. 대역폭을 고려한 측면값

모음의 제 1 포먼트와 제 2 포먼트 주파수의 대역폭이 각각 50Hz, 100Hz 정도임을 고려하여, 포먼트 주파수가 속하는 좌우 비트의 값을 0.0에서 0.4까지 0.1씩 증가시키며, 이와 동시에 초기 가중치와 학습대상을 변화시키면서 얻은 각각 7개의 인식율, 반복회수, 에러감소율을 평균하여 표 3에 나타내었다. 학습대상 패턴수가 60개, 측면값이 0.2인 경우를 제외하고는 측면값이 커질수록 대체로 수렴속도는 빨라지며 인식율은 높아짐을 알 수 있었다. 이는 종래의 방법보다 측면값을 고려하므로 인하여 입력정보가 많아졌기 때문으로 생각된다.

표 3. 측면값의 변화에 따른 수렴 속도와 학습 성취도의 관계

Table 3. The speed of convergence and the possibility of correct recognition, as the side value is varied.

측면값	학습과정시 계산된 패턴의 수								
	80개			60개			40개		
	인식율	반복 횟수	에러 감소율	인식율	반복 횟수	에러 감소율	인식율	반복 횟수	에러 감소율
0.0	99.2	501.6	-	95.1	505.7	-	81.1	476.7	-
0.1	99.5	483.0	37.5%	95.3	475.3	4.1%	83.8	464.7	14.3%
0.2	99.3	468.3	12.5%	94.5	451.2	-12.2%	83.6	464.1	13.2%
0.3	99.9	452.3	87.5%	96.6	441.5	30.6%	85.3	491.9	22.2%
0.4	99.9	448.4	87.5%	96.9	443.2	36.7%	90.1	469.3	34.9%

그림 6은 표 3의 측면값이 0.0, 0.2, 그리고 0.4인 경우에 대해 수렴 속도와 학습 성취도의 관계를 나타내었다. 여기서 수렴속도는 t_{ss} 값이 0.05이하로 감소될 때까지 소요된 총 시간을 분(min.)으로 나타내었다.

100개의 샘플중 80개를 택하여 학습시킨 후 100개의 샘플에 대한 인식 정도는 모두 99% 이상이였으며, 60개, 40개를 택하여 학습을 시켰을 때 100개의

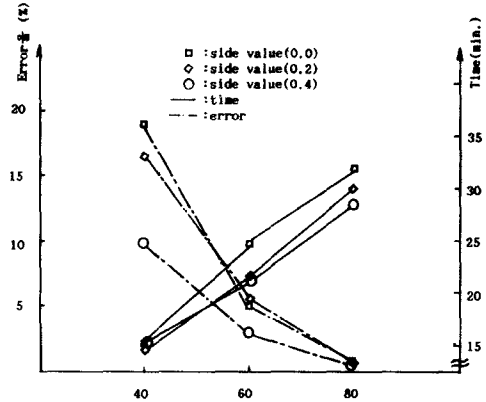


그림 6. 측면값의 변화에 따른 수렴 속도와 학습 성취도의 관계

Fig. 6. The speed of convergence and the possibility of incorrect recognition, as the side value is varied.

샘플에 대한 인식정도는 양측에 “0.0”을 두고 코오딩했을 때보다 본 연구에서 제안한 방법중 측면값을 “0.4”를 두고 코오딩했을 때 에러 발생율이 30% 이상 감소하며, 수렴 속도도 약 7% 이상 빠름을 알 수 있었다.

표 4는 오인식이 일어나는 샘플과 오인식된 결과를 알아보기 위하여, 먼저 신경망을 초기화 시켰으며 측면값을 “0.4”로 두고 코오딩하는 방법으로 총 학습용 샘플 100개중 40개만 학습시킨 후, 100개의 샘플을 입력케 인가했을 때의 인식 결과 분포이다. 여기서 샘플수는 총 학습용 샘플 100개 중 각 모음의 갯수이며, 기타는 바르게 인식하지 못한 갯수이다.

바르게 인식하지 못한 입력 패턴은 학습과정 중에 학습시키지 않은 입력 패턴 중에서 발생하였으며, 이들 에러는 주로 다음의 두가지 경우로 분류할 수 있다. 즉, 입력 패턴을 신경망에 인가하여 인식결과를 평가할 때, 출력층의 유니트들 가운데 바라는 출력 유니트의 활성값이 목표치(“1.0”)의 70%~90% 정도 되는 경우와 두개 이상의 출력 유니트에서 동시에 목표치의 70% 이상의 값을 나타내는 경우이다. 특히 전자의 경우에서 가장 큰 출력값을 나타내는 유니트의 모음을 입력에 인가한 모음으로 인식할 경우 전체 인식율을 93%에서 96%로 높일 수 있었다. 그리고 표 4에서 오인식의 대부분은 /o/인 경우에서 나타났으며, 이는 모음 /o/의 갯수가 다른 모음의 갯수에 비하여 적으며, 학습시 선택되는 샘플이 /o/의

표 4. 모음에 대한 인식 결과 분포

Table 4. The result of recognition for vowels.

입력	출력 샘플수	/a/	/e/	/i/	/o/	/u/	기타
/a/	24	24					
/e/	31		31				
/i/	12			12			
/o/	10				5		5
/u/	23					21	2

범위 중 일부분에 편중되어 선택된 것으로 생각된다. 따라서 인식율을 증가시키기 위해서는 어느 정도 이상의 갯수는 학습시켜주어야 하며 학습대상이 편중되지 않도록 해야 될 것으로 생각한다.

IV. 결 론

본 연구에서는 모음 인식 신경망을 학습시키기 위한 입력패턴을 구성할 때 포만트의 대역폭을 고려하여 측면값을 준 경우가 학습 성취도에 어떤 영향을 미치는가를 알아보았다. 본 연구에서 제안한 방법은 종래의 방법보다 에러율에서 30% 이상 감소하며, 수렴 속도면에서는 7% 이상 빠름을 시뮬레이션을 통하여 알 수 있었으며 주파수를 근거로 만든 패턴을 입력 패턴으로 사용하는 인식 시스템에서 수렴 속도와 학습 성취도를 향상시키는 방법으로 사용될 수 있음을 입증할 수 있었다. 기존의 pattern classification 방법등에도 본 연구에서 제안한 방법을 도입한다면 더 좋은 인식율을 얻을 수 있을 것으로 생각된다.

또한 본 연구의 학습 결과를 이용하여 실제 시스템을 구성한 후 음성 신호의 주파수 분석을 통하여 얻어진 제1, 제2 포만트 주파수를 인가할 경우 실시간적으로 출력층에서 모음 인식을 할 수 있을 것으로 생각된다.

參 考 文 獻

- [1] L.R. Rabiner and B.H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4-16, 1986.
- [2] T.Kohonen, "The neural Phonetic type-writer," *IEEE Computer*, pp.11-22, Mar.1988
- [3] B. Gold, "Hopfield model applied to vowel and consonant discrimination," *MIT Lincoln Laboratory Technical Report*, TR-747, AD-A169942, June 1986.
- [4] R.P. Lippmann and B. Gold, "Neural-net classifiers useful for speech recognition" *Proc. IEEE Int. Conf. Neural Networks*, pp. IV417-425, June 1987.
- [5] T. Ohotomo, K. Takahasi, and K. Hara, "Learning and discrimination of perceptual vowel distribution by a neural net model,"
- [6] 이황수, "신경회로 컴퓨터: 이론, 응용 및 구현 (음성인식)," 한국과학기술원 산학협동강좌, 1988.
- [7] 이기영, 이인섭, 최승호, 김자룡, 배철수, 최갑석, "Hopfield Network를 이용한 단모음 인식에 관한 연구," 대한전자공학회 하계종합학술대회논문집, 제12권 제1호, pp. 632-634, 1989.
- [8] 심재형, 이종혁, 윤태훈, 김재창, 이양성, "신경망을 이용한 모음의 인식 및 방법," 대한전자공학회 추계종합학술대회논문집, 제12권, 제2호, pp. 278-281, 1989.
- [9] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in D. E. Rumelhart and J.L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1: Foundations, MIT Press, pp 318-362, 1986.
- [10] B.A. Huberman, "An improved three layer back propagation algorithm," *Proc. IEEE Int. Conf. Neural Networks*, vol. 1, pp. 11637-643, July 1987.
- [11] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, pp. 74-75, 1978.

著 者 紹 介



沈 在 亨 (準會員)
 1965年 12月 20日生. 1988年 2月
 부산대학교 전자공학과 졸업. 1990
 年 2月 부산대학교 대학원 전자
 공학과 석사학위 취득. 1990年 2
 月~현재 효성컴퓨터 사업부 전
 자연구소 연구원. 주관심분야는
 컴퓨터 H/W, Neural net 등임.



李 鍾 赫 (正會員)
 1955年 5月 2日生. 1975年 2月
 부산대학교 전자공학과 졸업. 1980
 年 2月 부산대학교 대학원 전자
 공학과 석사학위 취득. 1990年 2
 月 부산대학교 대학원 전자공학
 과 박사과정 수료. 1980年 3月
 ~1990年 2月 동의공업전문대학 전자과 부교수. 1990
 年 3月~현재 경성대학교 컴퓨터공학과 조교수. 주관
 심분야는 컴퓨터 H/W, Neural net, 음성인식 등임.

尹 台 燾 (正會員) 第26卷 第10號 參照
 현재 부산대학교 전자공학과
 조교수



金 在 昌 (正會員) 第26卷 第10號 參照
 현재 부산대학교 전자공학과
 교수



李 梁 成 (正會員) 第26卷 第1號 參照
 현재 부산대학교 전자공학과
 교수