

# 음성인식을 위한 분산개념을 자율조직하는 신경회로망시스템

## (A Neural Net System Self-Organizing the Distributed Concepts for Speech Recognition)

金 成 錫\*, 李 泰 鎬\*

(Sung Suk Kim and Tai Ho Lee)

### 要 約

본 연구에서는 자기지도 BP 신경회로망의 은닉노드상의 활성패턴을 음성패턴의 분산표현된 개념으로 설정하고, 이 분산개념을 T.Kohonen의 자율조직 신경회로망(SOFM)의 입력특징으로 하는 복합회로망을 제안한다. 이렇게 함으로써 통상의 BP 신경망의 교육에 관련된 어려움과 패턴정합기로 떨어지는 약점을 해소하는 동시에 의미있고 다양한 내부표현을 추출해 낼 수 있다는 강점을 활용할 수 있고, SOFM의 강력한 판단기능을 이용하여 보다 구조적이고 의미있는 개념맵의 배열을 얻을 수 있게 되었다. 결과적으로 전처리가 불필요하고 자기교육이 가능한 독자적인 인식시스템이 구성된다.

### Abstract

In this paper, we propose a neural net system for speech recognition, which is composed of two neural networks. Firstly the self-supervised BP (Back Propagation) network generates the distributed concept corresponding to the activity pattern in the hidden units. And then the self-organizing neural network forms a concept map which directly displays the similarity relations between concepts.

By doing the above, the difficulty in learning the conventional BP network is solved and the weak side of BP falling into a pattern matcher is gone, while the strong point of generating the various internal representations is used. And we have obtained the concept map which is more orderly than the Kohonen's SOFM.

The proposed neural net system needs not any special preprocessing and has a self-learning ability.

\*正會員, 蔚山大學校 電子 및 電算機工學科  
(Dept. of Elec. and Computer Eng., Ulsan Univ.)  
接受日字: 1989年 3月 2日

I. 서 론  
신경회로망은 생물학적 두뇌의 구조에 기초한 인  
지적 정보처리 구조로서 연상기억, 학습, 패턴인식,

비전 및 음성에 유용함이 밝혀지고 있다.<sup>[11]</sup>

1970년대 후반부터 기존의 인공지능(A.I.) 기법의 문제점이 들어나고 신경회로망의 유용한 특성이 발견되면서 신경회로 컴퓨터는 각광을 받게 되었다.<sup>[12]</sup>

인간의 뇌에는 이미 우수한 음성인식기가 구현되어 있다는 관점에서 음성인식 문제에 있어 새로운 방법이 제시되고 있으며, T.Kohonen은 자발적으로 학습기능을 수행하는 자율조직 신경회로망(SOFM)을 고안하여 음소를 인식의 기본단위로 하는 음성인식 시스템을 개발하였고,<sup>[3,4,5,6,7]</sup> Elman, Gold, Bengio 등은 BP(back propagation) 신경회로망, 홉필드모델, 볼츠만 머신(BM) 등을 이용한 음성인식을 시도하였다.<sup>[8,9,10,11]</sup>

신경회로에서 지식(개념)을 저장하는 방법에는 국부적 방법(Barlow, 1972; Feldman, 1986)과 분산적 방법(Hinton & Rumelhart, 1986)이 있다.<sup>[12,13,17]</sup> 국부적 방법은 한개의 뉴론이 한개의 실체(entity)를 나타내는 방식(winner-take-all)이고 분산적 방법은 한개의 실체를 많은 뉴론상의 활성패턴(activity pattern)으로 나타내는 방식이다. 분산표현에서, 지식은 각 뉴론에 저장되는 것이 아니라 각 뉴론 사이의 시냅스 연결강도(weights)로 저장되며 새로운 지식을 저장한다는 것은 전에 저장되었던 지식에 영향을 주지 않고 시냅스의 연결강도를 조금씩 바꾸어 주는 것이다.

본 연구에서는 자기지도(self-supervised) BP 신경회로망의 은닉노드(hidden node)상의 활성패턴을 음성패턴의 분산표현된 개념으로 설정하고, T.Kohonen의 자율학습 알고리즘을 이용하여 개념이 국부적으로 표현되는 자율조직개념맵을 형성하였다. 이 자율조직개념맵은 개념패턴이 음성신호의 특징(feature)으로서 정보를 효과적으로 표현할 수 있음을 보이며, 개념간의 위상적 관계를 보존하고 있다. 물리적으로 유사한 개념은 개념맵상에서 인접한 뉴론에 배치된다. 제안된 신경회로망시스템은 인간의 1) 대뇌피질

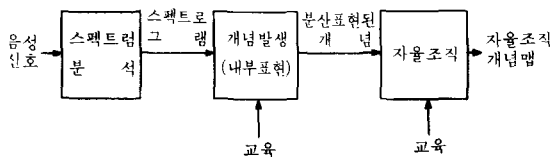


그림 1. 전체시스템의 구성도  
Fig. 1. The system organization.

은 많은 맵들로 구성되어 있고, 2) 청각신경은 대뇌 피질에 도달되기 전에 중간단계(lowbrain, midbrain)를 거쳐 대뇌피질로 전달되며, 3) 중간단계는 독특한 내부표현을 가진다는 관점<sup>[11,13]</sup>에서 사람의 음성인식구조에 근접하다고 할 수 있으며, 전체시스템의 구성도는 그림 1과 같다.

## II. 신경회로망시스템

### 1. 자기지도 BP 신경회로망(self-supervised BP neural network)

다층 신경회로망은 피드포워드(feedforward)와 피드백워드(feedbackward)의 시냅스(synapse)로 연결되어 있고, 학습이론으로 BP 알고리즘(Rumelhart, Hinton, & Williams, 1986)이 널리 이용되고 있다.<sup>[14,15]</sup> BP 알고리즘은 다층 인식자(multi-layer perceptron)에 "Delta rule(Widrow & Hoff, 1960)"을 다층구조로 일반화한 것으로 자체로서 하나의 인식기구를 형성할 수 있다. 이때 중요한 고려사항은 교육패턴과 입력신호의 형태이다. 교육과정에 한조의 입력패턴/희망출력패턴이 사전에 구성되어야 하므로 제한된 범위의 패턴정합기가 가지는 문제, 즉 어떤 패턴이 가장 적절한가 하는 별도의 분석이 요구되며 새로운 요구에 적응하려면 새로운 한조의 패턴을 구성해야 한다. 따라서 입력패턴은 독특한 전처리를 필요로 할 수 있다.

본 연구에서는 희망출력패턴을 입력패턴으로 동일하게 둬으로써(self-supervised) BP 신경망의 교육에 관련된 어려움을 해소하고, 은닉노드상에서 형성되는 활성패턴(분산표현된 개념)이 음성신호의 특징으로서 정보를 효과적으로 표현할 수 있음을 제시한다.

회로망은 그림 2와 같이 뉴론과 비슷한 특성을 지닌 요소들로 조직된 3층 신경망으로 입력층 및 출력층은 3×16 배열의 노드, 은닉층은 6개 노드로 구성되어 있다. 은닉노드의 수는 시냅스 공간(weight space)을 분할 할 수 있는 분리성(separability), 즉 신경망의 분류능력(discriminating capability)을 나타내며 최적인 은닉노드수의 사용은 효과적인 학습을 위해 매우 바람직하다.

본 연구에서는 대수투영법(algebraic projection)에 기초하여 은닉노드의 수를 결정하였다.<sup>[11]</sup> 입력패턴은 음성신호의 스펙트로그램이 이용되며 자기지도의 결과로 얻어지는 것은 회로망의 출력패턴이 아니라 은닉노드상에서 형성되는 분산개념패턴이다.

BP 알고리즘<sup>[14,15]</sup>과 유사한 자기지도 BP 신경회로망의 학습알고리즘은 다음과 같다.

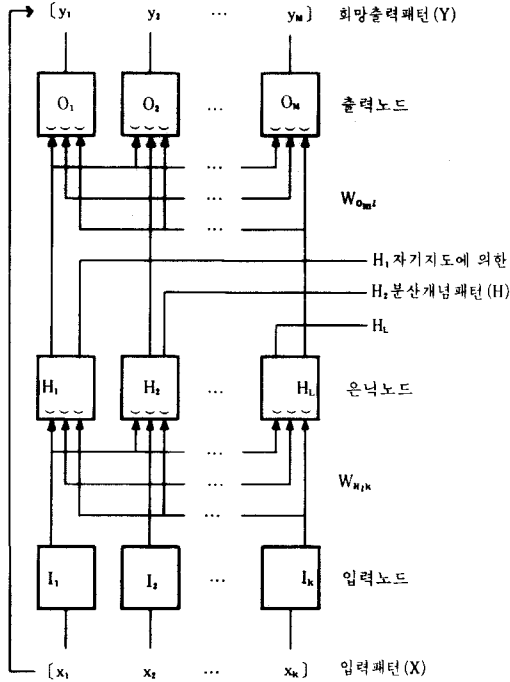


그림 2. 자기지도 BP 신경회로망  
Fig. 2. The self-supervised BP neural network.

1. 시냅스 연결강도(weights) 및 바이어스(bias)를 초기화 한다.
2. 입력노드 및 희망출력패턴에 랜덤잡음으로 변형된 입력패턴을 인가한다.

$$I = Y + r \times X, \quad (1)$$

$r$  :  $[-0.5, 0.5]$ 의 랜덤변수

3. 은닉노드의 출력값을 결정한다.

$$H_i(t+1) = \text{logistic}(\text{bias}_{H_i} + \sum_k W_{H_i,k} I_k(t)), \quad (2a)$$

$$\text{where } \text{logistic}(x) = 1 / (1 + \exp(-x/T)) \quad (2b)$$

4. 출력노드의 출력값을 결정한다.

$$O_m(t+1) = \text{logistic}(\text{bias}_{O_m} + \sum_i W_{O_m,i} H_i(t)) \quad (3)$$

5. 출력노드의 시냅스 연결강도 및 바이어스를 수정한다.

$$\Delta W_{O_m,i} = \eta \delta_{O_m} H_i + \alpha (W_{O_m,i}(t) - W_{O_m,i}(t-1)) \quad (4a)$$

$$\Delta \text{bias}_{O_m} = \eta \delta_{O_m} \quad (4b)$$

$$\delta_{O_m} = (y_m - O_m) O_m (1 - O_m) \quad (4c)$$

$\eta$  : 학습율 (learning rate),

$\alpha$  : 모멘텀 (momentum)

6. 은닉노드의 시냅스 연결강도 및 바이어스를 수정한다.

$$\Delta W_{H_i,k} = \eta \delta_{H_i} I_k + \alpha (W_{H_i,k}(t) - W_{H_i,k}(t-1)) \quad (5a)$$

$$\Delta \text{bias}_{H_i} = \eta \delta_{H_i} \quad (5b)$$

$$\delta_{H_i} = H_i (1 - H_i) \sum_m \delta_{O_m} W_{O_m,i} \quad (5c)$$

2. 자율조직개념맵 (self-organizing concept map)

T.Kohonen의 SOFM은 다음과 같은 근거를 가지고 제안된다. 1) 인간의 두뇌에서 감각경로 (sensory pathway)가 조직되는 중요한 원리는 뉴론의 배열이 규칙적이고 외부자극 (external stimulus)의 물리적 특성이 배열에 반영되는 것이다. 2) 위상적으로 인접한 뉴론은 물리적으로 유사한 자극에 민감하도록 조직되며 두뇌는 자극의 특성에 따라 자발적으로 뉴론을 조직하는 자율조직능력을 가지고 있다. 그리고 3) 두뇌에서 감각정보는 이들간의 위상적 관계를 보존하는 2차원의 맵 (map)으로 표현되며, 대뇌피질은 많은 맵들로 조직되어 있다.<sup>[11],[18],[20]</sup>

T.Kohonen은 2차원 배열의 SOFM이 내부표현을 가진다고 주장하고 있다.<sup>[8],[7]</sup> 그러나 SOFM은 4~6층으로 조직되어 있는 두뇌의 중간단계에서 형성되는 내부표현은 아니다. 본 연구의 분산개념패턴은 중간단계에서 형성되는 내부표현으로 간주될 수 있으며, T.Kohonen이 주장하는 정보의 내부표현방식과는 다르다. 한편, SOFM의 입력패턴은 전처리에 의해서 구조적인 정보로 표현되어야 맵상에서 특징이 발생된다.

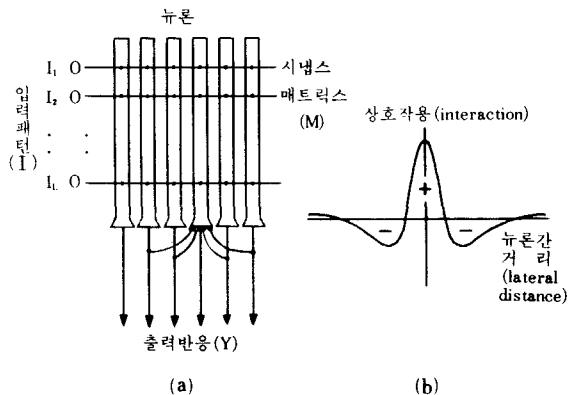


그림 3. (a) 자율조직 신경회로망  
(b) "Mexican-hat"형 상호작용  
Fig. 3. (a) The self-organizing neural network.  
(b) The "Mexican-hat" type interaction.

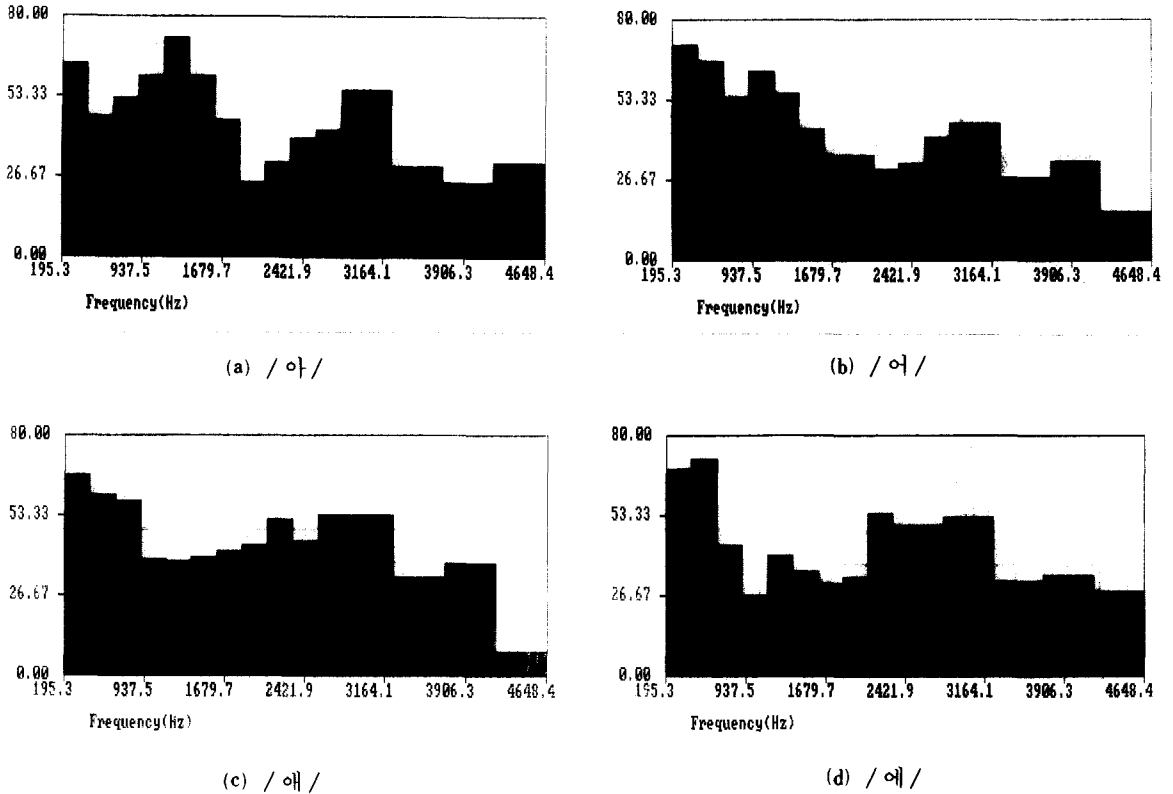


그림 4. 음성신호의 스펙트럼  
Fig. 4. Spectrums of speech signals.

본 연구의 자율조직개념맵은 전처리의 결과로 얻어진 것으로 간주될 수 있는 분산개념패턴을 입력패턴으로 하는 자율조직 신경회로망의 시냅스 매트릭스에 개념패턴이 적응변환(adaptive mapping)되어 형성되며, 물리적으로 유사한 개념은 맵상에서 인접한 뉴론에 배치된다. 자율조직개념맵을 형성하는 신경회로망은 그림 3과 같으며 뉴론의 특성과 비슷한 요소들로 구성된 8×8 배열로 조직되어 있다. 회로망의 요소들은 인접한 요소들(neighborhood)과 “Mexican-hat”형 측면피환(lateral feedback)함수로 서로 연결되어 있다. 그림에서 속이 빈 원은 자극적(excitation), 속이 찬 원은 억제적(inhibition) 영향을 이웃으로부터 받으며, 그림은 일차원이나 실제로는 이차원으로 구성되어 있다.

신경회로망의 학습이론은 외부지도(teacher)가 필요한 지도학습(supervised learning)과 자발적으로 학습하는 자율학습(unsupervised learning)으로 크게 분

류될 수 있다.<sup>[16,18]</sup>

본 연구의 자율학습에서 적응프로세스는 개념의 특성에 따라 학습과정에 시냅스 연결강도를 조절하여 시냅스 매트릭스(M)에 개념맵을 형성시킨다. 시냅스계(synaptic system)는 다음 미분방정식으로 표현된다.

$$dM/dt = f(I, Y, M), \tag{6}$$

여기서 I는 입력패턴, Y는 출력반응, 그리고 M은 시냅스 매트릭스이다.

자율조직개념맵을 형성하는 학습알고리즘은 다음과 같다.

1. 시냅스 연결강도를 초기화한다. ( $\sum m_{ij} = 1$ )
2. 입력패턴을 입력노드에 인가한다.
3. 입력벡터, I(t)에 가장 가까운 뉴론 K를 선택한다.

$$\| I(t) - M_k(t) \| = \min \{ \| I(t) - M_i(t) \| \} \tag{7}$$

4. 뉴런 K와 위상적으로 이웃에 있는 뉴런들,  $N_k(t)$ 의 연결강도를 조절한다.

$$M_i(t+1) = M_i(t) + \alpha(t)[I_i(t) - M_i(t)] \quad (8a)$$

for  $i \in N_k(t)$

$$M_i(t+1) = M_i(t) \text{ for } i \notin N_k(t) \quad (8b)$$

$\alpha$ : 적응이득(adaptation gain)

뉴런 K에 이웃하는 뉴런들(neighborhood),  $N_k(t)$ 는 뉴런 K로 부터 일정한 반경에 있는 모든 뉴런들을 포함하고 있으며, 처음에는 회로망의 모든 뉴런을 포함하고 있으나 시간의 단조감소함수로 점차 줄어들어 마지막에는 뉴런 K만 포함한다.

적응이득,  $\alpha(t)$ 는 개념맵이 초기에는 거칠게, 이후에는 정교하게 조직되도록 두 단계의 단조감소함수로 적용된다.

$$\text{단계 1. } \alpha_1(t) = k_1(1-t/t_1) \text{ for } 0 \leq t \leq t_1 \quad (9a)$$

$$\text{단계 2. } \alpha_2(t) = k_2(1-t/t_2) \text{ for } t_1 < t \leq t_2 \quad (9b)$$

$k_1 > k_2, t_1 \ll t_2.$

### III. 실험 및 결과고찰

실험에 사용된 음성자료는 남녀 1인이 발음한 우리말의 모음(/아(a)/, /어(@)/, /오(o)/, /우(u)/, /으(w)/, /이(i)/, /애(e)/, /에(E)/)으로 3.4KHz의 저

역통과필터를 거쳐 10KHz로 샘플링한 12Bit의 디지털 데이터로 컴퓨터에 저장되었다.

음성신호의 분석은 6.4ms중첩된 128-샘플을 한 프레임으로 하고, 128-샘플 해밍윈도우를 이용하여 256-샘플 FFT를 수행하였다. 스펙트럼은 190Hz-4800Hz 사이의 16개 에너지 성분으로 구성되며(그림 4), 스펙트로그램은 25.6ms 동안의 3개 프레임으로 구성된다.

본 연구에서는 스펙트로그램을 평균값으로 정규화한 다음 logistic함수  $[1/(1+\exp(-x/T))]$ 를 이용하여  $[0, 1]$ 사이의 값으로 변환한 후에 랜덤잡음을 첨가하여 자기지도 BP 신경회로망의 입력패턴으로 하였고, 또한 BP 신경회로망에서 발생한 분산개념패턴을 평균값으로 정규화( $\sum H_i = 1$ )한 다음 랜덤잡음을 혼합하여 자율조직 신경회로망의 입력패턴으로 구성하였다.

자기지도 BP 신경망의 원 음성신호에 대한 분산개념패턴은 그림 5(a), 그리고 신호대잡음비, SNR=2.4로 변질된 음성신호에 대한 분산개념패턴은 그림 5(b)와 같다. 이 개념패턴들은 100,000 교육사이클 후에 신경망의 은닉노드상에서 형성된 활성화패턴이며 그림에서 사각형의 면적은 은닉노드의 활성화레벨(activation level)을 나타낸다.

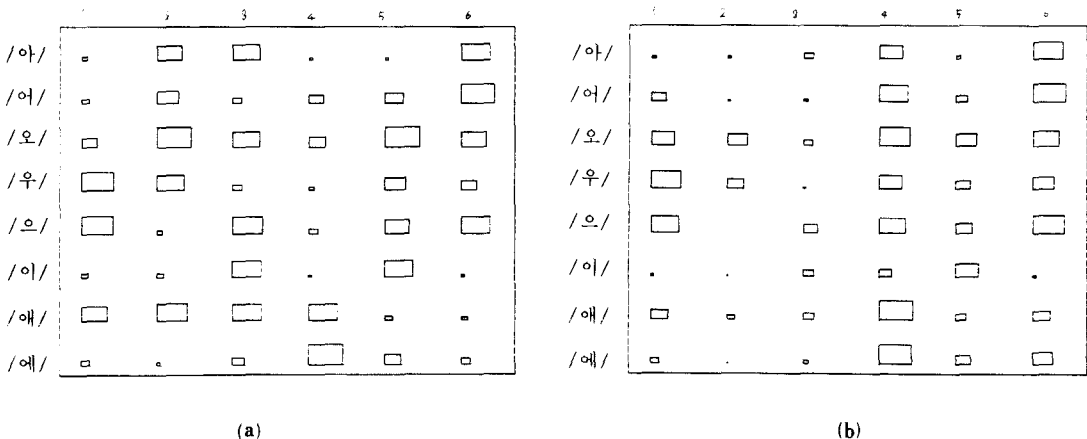


그림 5. (a) 원 음성신호의 개념패턴  
(b) 변질된 음성신호의 개념패턴

Fig. 5. (a) The concept patterns of raw speech signals.  
(b) The concept patterns of corrupted speech signals.

자율조직 신경회로망에서 개념패턴을 입력특징으로 하여 형성된 본 연구의 자율조직개념맵은 그림 6(a), 그리고 스펙트럼 정보를 입력특징으로 하여 형성된 T.Kohonen의 SOFM은 그림 7과 같다. 이 그림으로 부터 자율조직개념맵상의 clustering이 T.Kohonen의 SOFM보다 우수함의 정도를 정량적으로 평가할 수는 없으나 개념패턴이 더욱 구조적이고 정보를 압축하고 있음을 알 수 있다. 그림 6(b)에서 원으로 표현된 요소는 그림 5(a)의 원 음성신호의 개념패턴에 대한 반응이고 사각형으로 표현된 요소는 그림 5(b)의 변질된 음성신호의 개념패턴에 대한 반응을 나타낸다.

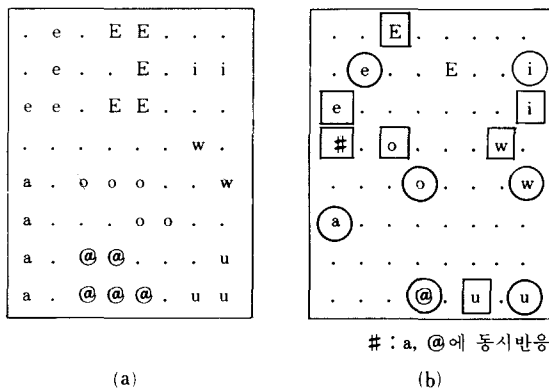


그림 6. (a) 자율조직개념맵  
(b) 개념맵 상에서 반응  
Fig. 6. (a) The self-organizing concept map.  
(b) The responses in self-organizing concept map.

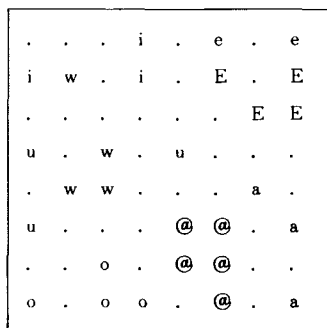


그림 7. T.Kohonen의 SOFM  
Fig. 7. T.Kohonen's self-organizing feature map.

본 논문에서 제안된 신경회로망시스템의 연구결과를 다음과 같이 요약할 수 있다.

1. 자기지도 BP 신경회로망은 교육상의 어려움을 해소하고, 자기지도에 의해 형성되는 은닉노드상의 활성패턴은 입력신호의 정보를 압축, 분산표현한다.
2. 분산표현된 개념패턴의 정보는 구조적이다.
3. 신호에 비례하는 랜덤잡음을 입력패턴에 혼합하면 시스템의 성능이 개선된다.
4. 인간의 두뇌는 중간단계에서 내부표현을 가진다는 관점에서 제안한 신경회로망시스템은 사람의 음성인식구조에 근접하다고 할 수 있다.
5. 개념맵에서 발생한 결정이 개념패턴의 발생에 간여할 수 있는 귀환루프를 둬으로써 좀더 사람의 인식구조에 유사한 동적(dynamic) 신경회로망시스템을 구성할 수 있을 것으로 사료된다.

参 考 文 献

- [1] P.K. Simpson, "A survey of artificial neural system," San Diego Engineering Center, UCSD, CA 92117, 1987.
- [2] S.E. Fahlman and G.E. Hinton, "Connectionist architecture for artificial intelligence," *IEEE Computer*, pp. 4-22, Jan. 1987.
- [3] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.* 43, pp. 59-69, 1982.
- [4] T. Kohonen, K. Makisara and T. Saramaki, "Phonotopic maps-Insightful representation of Phonological features for speech recognition," *IEEE 7th Inter. Conf. on Pattern Recognition*, pp. 182-185, 1984.
- [5] T. Kohonen and K. Makisara, "Representation of sensory information in self-organizing feature maps," in J.S. Denker (Ed.) *AIP Conference Proceedings 151, Neural Networks for Computing*, Snowbird Utah, AIP, 1986.
- [6] T. Kohonen, "Self-organization and associative memory," Springer-Verlag, New York, 2nd ed. 1988.
- [7] T. Kohonen, "The neural phonetic typewriter," *IEEE Computer*, pp.11-22, Mar. 1988.
- [8] J.L. Elman and D. Zipser, "Learning the hidden structure of speech," Institute for Cognitive Science, UCSD, ICS Report 8701, Feb. 1987.