SOFM 신경회로망을 이용한 한국어 음소 인식

Korean Phoneme Recognition Using Self-Organizing Feature Map

전 용 구*, 양 진 우*, 김 순 협*
(Yong-Koo Jeon*, Jin-Woo Yang*, Soon-Hyob Kim*)

요 약

본 논문에서는 패턴 매칭 방법에 근거하여 인식 단위가 음소인 음소 기반 인식 시스템을 구성하였다. 선택한 신경망 구조는 생물학적 신경망인 코호넨(T. Kohonen)의 SOFM(Self-Organizing Feature Map)으로 패턴 매칭 과정 중 클러스터러(clusterer)로 사용하였다. SOFM 신경망은 신호 공간에 대해서 최적의 국소(局所) 해부적 사상(local topographical mapping)에 의한 자기 조직화 과정을 수행하며, 그 결과 인식 문제에 있어서 상당히 높은 정확도를 나타낸다. 따라서 SOFM 신경망은 음소 인식에도 효과적으로 응용될 수 있다. 또한 음소 인식 시스템의 성능 향상을 위해 K-means 클러스터링 알고리즘이 결합된 학습 알고리즘을 제안하였다.

제안된 음소 인식 시스템의 성능을 평가하기 위해 먼저, 인식 대상 음소는 모음군 17개, 자음의 경우 파열음 9개, 마찰음 3개, 파찰음 3개, 유음 및 비음 4개, 음소의 성질이 다른 종성 7개의 음소군으로 모두 43개의 음소를 대상으로 실험하였으며, 각 음소군에 대한 특징 지도를 구성하여 레이블러(labeler)의 기능을 수행하게 하였다. 화자 종속 인식 실험 결과 87.2%의 인식률을 보였으며 제안한 학습법의 빠른 수렴성과 인식률 향상을 확인하였다.

ABSTRACT

In order to construct a feature map-based phoneme classification system for speech recognition, two procedures are usually required. One is *clustering* and the other is *labeling*. In this paper, we present a phoneme classification system based on the Kohonen's *Self-Organizing Feature Map*(*SOFM*) for clusterer and labeler. It is known that the SOFM performs self-organizing process by which optimal local topographical mapping of the signal space and yields a reasonably high accuracy in recognition tasks. Consequently, SOFM can effectively be applied to the recognition of phonemes. Besides to improve the performance of the phoneme classification system, we propose the learning algorithm combined with the classical *K-means clustering algorithm* in fine-tuning stage.

In order to evaluate the performance of the proposed phoneme classification algorithm, we first use totaly 43 phonemes which construct six intra-class feature maps for six different phoneme classes. From the *speaker-dependent* phoneme classification tests using these six feature maps, we obtain recognition rate of 87.2% and confirm that the proposed algorithm is an efficient method for improvement of recognition performance and convergence speed.

*광운대학교 전자계산기공학과
Dept. of Computer Engineering, Kwang Woon Univ.
접수일자 : 1995년 1월 16일

I. 서 론

음성 인식의 방법은 주로 패턴 정합(pattern matching) 방법에 근거를 두고 있으며, 최근에는 애널로그, 디지틀 VLSI 제조 기술의 발전으로 병렬 처리가 가능해짐으로써 인간의 두뇌를 모방한 신경 회로망(Neural Network)을 이용한 음성 인식 기술이 대두되고 있다[15].

한편, 음성 인식의 최종 목표는 (1)불특정 화자(speaker independence) (2)연속 음성 (3)대어휘의 실현으로 볼 수 있는데 대어휘 연속 음성 인식 시스템에서는 어휘 수가 많아짐에 따라 기억해야 할 데이타의 수가 많아지고 인식시 계산량이 많아지는 문제를 피하기 위해 단어 단위의 모델링(modeling)을 피하고 변이음(allophone), 음소(phoneme), 다이폰(diphone), 음절(syllable) 등의 단어 하부 단위(sub-word unit)로 모델링 하게 된다. 특히 형태소의 결합 유형에서 볼 때 (W. von Humboldt) 교착어(agglutinative language)에 속하는 한국어는 그 어형 성법에 있어서 굴절, 파생, 합성 세 가지 모두를 널리 사용하므로[12] 그 필요성이 더욱 요구된다고 하겠다.

본 논문에서는 음소(phoneme)를 인식 단위로 하는 음소 기반 인식 시스템을 구현함에 있어서 패턴 정합(pattern matching) 기법을 사용하였고 클러스터링(clustering)에 의한 표준 패턴(reference pattern) 생성시 기존의 반복적(iterative) 기법인 K-means 알고리즘의 제약을 해결하고자 클러스터러(clusterer)로 T. Kohonen의 SOFM(Self-Organizing Feature Map) 신경 회로망을 사용하였으며 clusterer의 성능 향상을 위해 K-means 학습을 결합하였다. 본 논문에서 구현한 시스템의 궁극적인 목적은 음성 신호(signal)를 음운학적 기호(symbol)인 유사-음소열(quasi-phoneme sequence)로 변환시키는 레이블러(labeler)의 역할을 수행하는 것이다.

II. 패턴 인식과 신경 회로망

패턴 인식에 관한 연구는 인간이나 동물의 패턴 인식 능력 자체를 탐구하는 분야와 특정한 응용을 위해 주어진 인식 작업이 가능한 인식기를 설계하는 이론 및 기술 개발 분야의 두 가지로 나뉠 수 있으며 공학적 관점에서의 패턴 인식 과정은 외계의 사상(event)을 관측, 해석함으로써 이미 자기 내부에 형성된 정보 모델에 따라 주어진 패턴에 분류 표시를 하여 그것이 속하는 부류(class/category)의 명칭을 출력하는 과정이라고 정의할 수 있다.

패턴 인식은 접근 방법에 있어서 전통적으로 중요한 확률적(statistical) 또는 결정 이론적(deterministic/decision theoretic) 방법과 구조 해석적(syntatic) 방법의 두 가지가 있어왔다. 최근에 각광받는 신경망(Neural Networks) 기술은 그 세 번째 접근 방법으로서 기존의 패턴 생성 메카니즘(mechanism)을 모델링하는 것을 피하고 대신에 입/출력으로부터 또는 블랙 박스의 관점에서 문제를 다루는 것이다. 즉, 인간의 두뇌는 입/출력 특성을 정량화하는 자세한 알고리즘 세트가 없이도 지능적인 동작(패턴 인식과 분류를 포함하는)을 관찰하고 흉내낼 수 있다는 사실로부터 좋은 블랙 박스 모델이 된다.

음성 인식을 위한 신경 회로망의 구조는 크게 구조상 지연(delay) 요소, 피드백(feedback) 연결이 없는 비회귀(nonrecurrent) 형태로 정적(static) 구조인 다층 인식자(MLP, Multilayer Perceptron)와 SOFM(Self-Organizing Feature Map)이 있으며, 전방향 연결선 외에 시간적으로 현재 입력에 대해 과거 시점에서의 상태를 반영해 신경 회로망의 출력을 결정하는 동적(dynamic) 구조가 있다.

III. K-means 알고리즘이 결합된 SOFM 신경 회로망

인간 두뇌의 정보 처리 과정을 살펴보면 특별한 교사 신호(teacher)없이도(unsupervised) 외계의 정보 신호에 따라서 그것을 뇌 속에 표현하는 메카니즘을 지니고 있다는 사실이다. 예를 들면 외계 신호 공간의 정보 구조를 2차원인 신경장(fields, layer, slab)으로 사상(mapping)하여 표현하는 자기 조직화(self-organization)과정의 결과로 정보를 국소(局所)적(local)으로 표현하는 것을 볼 수 있다. 이러한 자기 조직화 과정은 레이블링 되지 않은 데이타를 부분 집합으로 분할(partitioning)하는 넌파라메트릭(nonparametric) 접근법에 해당한다.

3.1 클러스터링 알고리즘

생물학적 신경망에서 수행되는 자기 조직화 과정은 클러스터링과 같은 개념으로 유클리드 공간에서 점들로 간주되는 패턴 벡터들 간에 유사도(similarity)

를 계산함으로써 그들 간의 근접도(proximity)에 따라 '군집(clusters)을 이루게 하는' 과정이다. 결과적으로 군집 영역은 서로 다른 패턴 부류로 해석될 수 있으며 주어진 패턴 세트의 부류를 결정하는 것은 결국 해당되는 군집을 찾는 것으로 귀결된다. 이러한 클러스터링 특성은 거리 개념을 기초로 한 분류기(distance-based pattern classification)의 성능에 중요한 역할을 한다.

3.2 유사도 계산(similarity measure)

일반적으로 다음과 같은 계산이 필요하다.

$d(x_i, x_j)$

$$= \begin{cases} \text{'크다'} & x_i\text{와 } x_j \text{ 가 다른 클러스터에 속한 경우} \\ \text{'작다'} & x_i\text{와 } x_j \text{ 가 같은 클러스터에 속한 경우} \end{cases} \tag{3.1}$$

3.3 K-means 알고리즘

패턴 유사도 계산이 선택되면 주어진 데이타를 클러스터 영역들로 분할하는 절차를 구체화해야 하는 문제에 직면하게 된다. 클러스터링 알고리즘은 크게 반복적(iterative) 접근법과 계층적(hierachical) 접근법으로 분류되며 계층적 접근법은 클러스터의 합병(merging) 또는 클러스터의 분리(splitting)로 세분화된다.

계층적 전략은 데이타 분할에 있어서 모든 경우를 고려하지 않는 특성을 가지므로 탐색 계산량이 상당히 감소되는 점에서 매력적이지만 표본의 수가 커지면 부적절한 것으로 알려져있다. 반복적 접근법의 경우 클러스터링 평가 함수(criterion function) 또는 성능 지수(performance index)의 최소화/최대화에 기초하여 스스로 반복적인 절차를 통해 주어진 데이타에 내재하는(underlying) 자연적인 그룹들(natural groups/clusters)을 발견하게 된다. 이러한 접근법의 대표적인 것이 K-means 알고리즘이다. 여기서 K란 단순히 부류의 갯수를 가리킨다.

평가 함수의 역할이 중요함을 먼저 살펴보자. n개의 벡터를 K개의 부분 집합으로 다음과 같이 분할 가능하다.

$$\frac{1}{K!}\sum_{p=1}^{K} {}_KC_p(-1)^{K-p}P^n \approx \frac{K^n}{K!} \tag{3.2}$$

예를 들면, n=100의 벡터들과 K=5인 경우 근사적으로 $5^{100}/5! \approx 10^{68}$의 분할이 가능하다. 이러한 탐색은 명백히 비실제적이므로 효율적인 계산 방식으로 분할(partition), P를 찾아야 할 필요가 있음을 알 수 있다. 따라서 다음을 만족하는 클러스터링 평가 함수 J(P)를 도입한다.

$$J(P_{min}) = P_{min}(J(P)) \tag{3.3}$$

가장 일반적으로 사용되는 평가 함수 J(P)의 정의는 다음과 같다. 주어진 훈련 세트(S_U)에 N개의 표본(x)이 존재할때 표본 평균(mean/centroid) 벡터를 m이라 하면

$$m_p = \frac{1}{N_p}\sum_{x\in S_p} x \tag{3.4}$$

이고, 평가 함수 J(P)를 오차의 제곱-합(Sum of Squared Error, SSE)으로 정의 한다.

$$J_{SSE}(P) = \sum_{p=1}^{K}\sum_{x\in S_p}\|x-m_p\|^2 \tag{3.5}$$

따라서 J_{SSE}는 주어진 분할에 대해서 전체 '분산(variance)'을 의미한다. 여기서 집합 S_U의 분할, P는 S_U의 분리된 부분 집합들로 이루어진 집합이다. 즉 $S_i = P = \{S_1, S_2, ..., S_K\}$ 이다.

K-means 알고리즘의 절차는 다음과 같다.

1. K개의 초기 클러스터 중심 $m_1(1), m_2(1), ..., m_K(1)$을 선택한다. 이 들은 주어진 표본 집합의 첫 번째 K개 표본으로, 임의로 선택된다.
2. k 번째 반복 스텝에서 다음의 관계를 이용하여 K 클러스터 영역내의 표본 {x}를 분배한다.

$$x \in S_j(k) \text{ if } \|x-m_j(k)\| < \|x-m_i(k)\| \tag{3.6}$$

모든 i=1, 2, ..., K에 대해서 $i \neq j$이고 $S_j(k)$는 클러스터 중심이 $m_j(k)$인 표본들의 집합을 가리킨다. 여기서 유사도 계산은 유클리드 거리 ‖.‖이다.

3. 스텝 2의 결과로 부터 $S_j(k)$의 모든 점들로부터 새로운 클러스터 중심까지의 거리 제곱-합이 최소가 되는 새로운 클러스터 중심 $m_j(k+1)$을 계산한다. 여기서 j=1, 2, ..., K 이다. 바꿔 말하면 새로운 클러스터 중심 $m_j(k+1)$은 평가

함수의 최소화를 만족한다.

$$J_j = \sum_{x \in S_j(k)} \| x - m_j(k+1) \|^2 \qquad (3.7)$$

여기서, $m_j(k+1)$은 단순히 표본 평균(mean)이며, 새로운 클러스터 중심은 다음과 같이 주어진다.

$$m_j(k) = \frac{1}{N_j} \sum_{x \in S_j(k)} x, \quad j = 1, 2, \ldots, K \qquad (3.8)$$

여기서, N_j는 $S_j(k)$안의 표본 수이다.

4. 만일 $m_j(k+1) = m_j(k)$, $j = 1, 2, \ldots, K$이면 알고리즘은 수렴(convergence)되었고 종료한다. 그렇지 않으면 스텝 2로 간다.

한편, K-means 알고리즘은 다음 네 가지의 제약을 지니고있다.

(1) 명시되는 클러스터 중심의 갯수, K

(2) 초기 클러스터 중심의 선택

(3) 표본이 취해지는 순서

(4) 주어진 데이타의 통계학적 분포(확률 분포 함수) 특성

여기서 (1)과(2)는 '클러스터 유효성(validity)'의 연구 분야로서 흥미있는 주제가 되고 있다[1]. 본 논문에서는 이러한 네 가지 제약성을 SOFM 신경 회로망[3]을 통해 해결하고자 한다.

3.4 SOFM(Self-Organizing Feature Map) 신경회로망

인식 세포 혹은 특징 추출 세포가 자기 조직에 의하여 형성되는 것을 신경 회로망 모델로서 유도하여 컴퓨터 시뮬레이션(simulation)으로 이러한 자기 조직이 가능함을 제시한 것은 von der Malsberg(1973)였다[4]. 그 후 T. Kohonen(1987)[5]은 보다 단순화된 모델을 기초로 대규모 컴퓨터 시뮬레이션을 하였다. Grossberg의 적응 공명 이론(Adaptive Resonance Theory)이나 Fukushima의 Cognitron, Neocognitron도 같은 자기 조직 기능을 기반으로하여 학습 능력을 갖는 실용적인 패턴 인식 장치를 만들려고 하는 정밀한 모델이다[14]. 이제 외계 신호 공간에 있어서 신호의 상관 관계를 표현하는 위상(topology) SOFM 신경망은 입력 정보를 2차원의 신경장에 국소화(localization)시켜 응답하게 하는 메카니즘인 경쟁 학습을 통해 입력 패턴 내에 존재하는 어떤 구조를 발견하는 특징 검출기의 역할을 한다. 본 논문에서 사용한 SOFM 신경 회로망의 구조는 그림 3.1과 같은 2차원의 평면 위상(planar topology)을 형성하는 노드의 연결 구조를 갖는다.

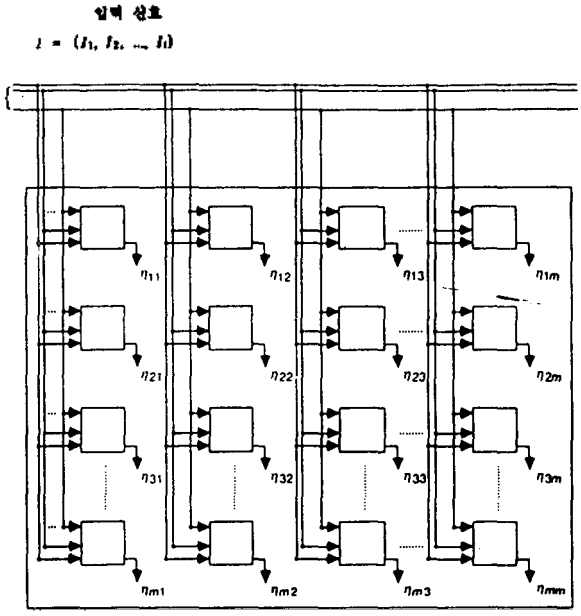


그림 3.1 2차원 평면 위상 형태의 SOFM 신경 회로망
Fig. 3.1 A two-dimensional planar topology of SOFM

물론 3차원과 그 이상의 차원을 갖는 위상도 가능하지만 P. Brauer 등의 연구[6]에 의하면 2차원 위상보다 성능이 떨어지는 것으로 보고되었다. 그 위상에 관계 없이 각 노드는 병렬로 입력 패턴 $I = (i_1, i_2, \ldots, i_d)$을 받아들인다. 선택된 노드들의 위상학적 배열(topological ordering)을 고려함으로써 d차원의 특징 공간(feature space)은 2차원의 특징 지도상에 사상된다. 즉 자기 조직화 특성이 갖는 장점인 차원 축소(dimensionality reduction)[7]가 이루어지며 축소된 차원 공간(constraint surface)에서 좌표축은 노드들간의 유사도 관계를 반영하며 지도상에서의 위상학적 거리는 비유사도(dissimilarity)에 비례한다. 선택된 지도의 차원은 또한 네트웍의 학습 시간에 영향을 주게된다. 그러나 효과적인 결과는 단지 1, 2 차원 위상을 사용함으로써 얻어진다는 사실에 유의할 필요가 있다.

본 논문에서 유사도 계산을 위해 유클리드 노름(Euclidean Norm)을 사용하였고, 유클리드 거리 계산을 위해 취하는 제곱근을 벗기면 다음과 같이 전개될 수 있다

$$d^2(I, W) = \|I-W\|^2$$
$$= [\|I\|^2 + \|W\|^2 - 2\langle I, W\rangle]$$
$$= 2(1-\cos\theta) \quad (3.9)$$

여기서, 입력 벡터와 연결 강도 벡터의 정규화의 근거를 발견할 수 있으며 $\cos\theta$ 항은 입력 벡터와 연결 강도 벡터의 상관(correlation)을 반영하며 정규화된 정합(normalized matching)을 나타낸다. 즉 cosine 거리[10]가 된다. 위의 식을 보면 오차의 제곱(Squared Error)으로 표현되는데 여기서 입력 신호를 통계적으로 정적(stationary)이라고 가정하면 입력 신호 집합에 대한 이들 값의 합은 평균값(average, mean value/expectation value)에 비례하므로 $d^2(I, W)$는 오차의 제곱 평균(Mean-Square-Error, MSE)이 되며 학습은 결국 이 값을 최소화하는 LMSE (Least-Mean-Square-Error) 문제와 일치함을 알 수 있다. 또한 이것은 입력 벡터 I를 목표(target) 출력으로 보았을 때의 Delta Rule(Widrow-Hoff Rule)과 동일하며 이 Delta Rule이 최소 제곱 평균(LMS) 또는 기울기 급강하법(Gradient-descent Rule)임을 기억하면 연결 강도 벡터 W는 그 자신과 입력 벡터 I사이의 오차를 그림 3.2[9]과 같이 가장 급한 기울기 방향으로 줄여 가는 것을 알 수 있다. 따라서 식 (3.9)는 네트웍의 연결 강도 수정을 위한 오차 함수(error function)로 해석되며, 학습은 연결 강도 값의 최적화(optimization/relaxation) 문제로 귀결 된다.

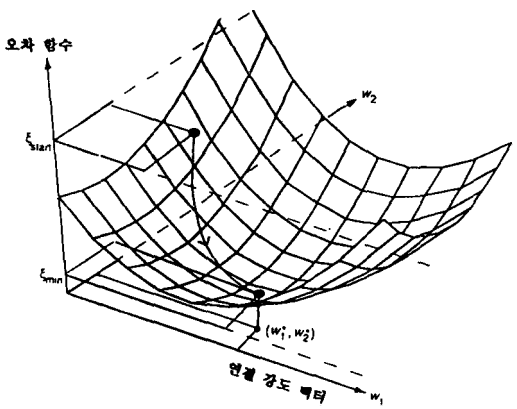


그림 3.2 기울기 급강하법에 의한 연결 강도 조정
Fig. 3.2 Visualization of the gradient-descent rule

위에서 언급한 바와 같이 SOFM 신경 회로망의 학습은 오차의 제곱 평균(MSE)을 최소화시키는 과정으로 해석할 수 있다. 그러나 실제로 학습 과정을 중지시키는 판단 기준은 학습률 계수 α 또는 반복 횟수가 되며 물론 이것은 네트웍이 충분히 훈련(학습)되었음을 보장해 주지 않는다. 예를 들면 학습 패턴 세트가 적절히 부류를 나타내지 못하는 경우 출력 층의 연결 강도 벡터는 정확한 중심점을 찾아내지 못하게 되므로 결국 충분한 훈련이 되지 않았음을 의미한다. 이러한 상황이 그림 3.3에 나타나있다. 즉 W_2가 부류 2의 올바른 중심점을 학습하지 못함으로 인해 패턴 x_2는 실제적으로 W_1에 더 가까우며 부류 1로 오분류(misclassification)된다. 따라서 특징 지도의 자기 조직화 특성(locality)을 보존하면서 연결 강도 벡터를 재조정(retuning)할 필요가 있다[11].

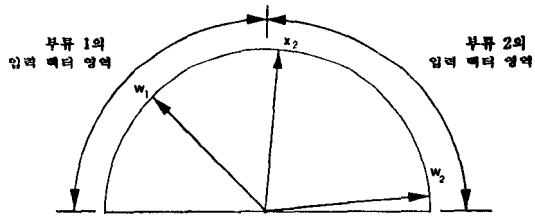


그림 3.3 국부 최소점 문제에서 기인하는 오분류의 예
Fig. 3.3 Example of misclassification resulted from local minima problem

본 논문에서는 K-means 클러스터링에서 도입했던 오차 제곱의 합(SSE) 평가 함수를 최소화하도록 특징 지도의 연결 강도 벡터를 미세 조정함으로써 국부 최소점(local minima) 문제를 해결 하고자하였다.

즉 특징 지도의 오차 제곱의 합 또는 전체 '분산(variance)'을 다음과 같이 정의 하고

$$V_{SSE} = \sum_{p=1}^{K} \sum_{I \in S_p} \|I - W_p\|^2 \quad (3.10)$$

V_{SSE}가 국부 최소점으로 부터 탈출할 수 있도록 다음의 K-means 방법을 이용하여 연결 강도를 재조정한다.

1. 학습 패턴 세트 S_p를 훈련된 특징 지도의 각 노드에 할당하고 분할된 각 S_p(p=1, 2, ..., K)에 대해서 유클리드 거리가 최소인 노드를 선택한다.
2. 선택된 각 노드의 연결 강도를 해당하는 분할 패

턴 세트 S_p의 중심점(centroid)으로 바꾼다.

$$W_p(i,j)=\frac{1}{N_p}\sum_{I\in S_p} I \qquad (3.11)$$

여기서, i, j 는 M*M인 특징 지도상의 노드 인덱스(index)이고

$1\le i, j\le M$ 이다.

3. $1\le p\le K$에 대해 스텝 1, 2를 반복한다.

위의 K-means 학습은 제곱 평균(MS, mean-square)의 의미에서 수렴(convergence)을 만족한다. 좀 더 엄격한 의미에서 수렴은 주어진 학습 절차가 사상(mapping)(여기서는 입력 과 연결 강도 벡터간의)을 적절히 포착(capture)하는 능력을 갖는지 분석하는 수단이 된다. 따라서 K-means 학습은 유용하다고 할 수 있다.

네트워의 전체적인 학습 절차는 그림 3.4와 같다.

```
Procedure FeatureMap_Learning( )

/* 출력 층 노드와 연결 강도 초기화 */
for ( i ← 1 ; i ≤ OutputNodesNumber ; i ← i+1 ) {
    for ( j ← 1 ; j ≤ InputNodesNumber ; j ← j+1 ) {
        W_ij ← random ( )
}

for ( ε > 0 && nl ≤ Iteration ) {

    /* 학습 패턴 벡터 입력 */
    for ( S_p ∈ { 학습 세트, S_U } ) {
        /* 거리 계산 */
        for ( i ← 1 ; i ≤ OutputNodesNumber ; i ← i+1 ) {
            distance (i) ← 0
            for ( j ← 1 ; j ≤ InputNodesNumber ; j ← j+1) {
                distance (i) ← distance (i) + d(I_j , W_ij)
        }

        /* 승리 뉴런 찾기 */
        k ← i
        for ( i ← 1 ; i ≤ OutputNodesNumber ; i ← i+1 ) {
            # ( distance (i) < distance (k) )
                k ← i
        }

        /* 연결 강도 조정 */
        # ( K-means Training )

            W_k ← (1/N_p) Σ_{I∈S_p} I

        else
            for ( Neighbor(k) ) {
                for ( j ← 1 ; j ≤ InputNodesNumber ; j ← j+1 ) {
                    W_kj ← W_kj + ε *( I_j - W_kj )
            }

        /* 파라미터 조정 */
        # ( ε ≥ T_segment )
            ε ← α_ε * ε
            Neighbor ( ) ← α_NB * Neighbor ( )
            T_segment ← α_T * T_segment

    }/* for S_P */
    # ( K-means Training )
      QUIT Iteration Loop

}/* Iteration */
```

그림 3.4 K-means 클러스터링이 결합된 SOFM 학습 절차
Fig. 3.4 Learning procedure of SOFM combined with K-means clustering

Ⅳ. 시뮬레이션 결과 및 고찰

4.1 우리말의 음소

한 언어에서 사용되는 모든 음성들이 언어적으로 유의미한 기능을 가지는 것은 아니며, 따라서 어떠한 단위가 언어적으로 유의미한 변별적 기능을 갖는가를 관찰할 필요가 있다. 여기서는 그 단위로 음소(phoneme)를 선택하였다. 음소 분석은 간단하지 않으며 어떤 방식으로 분석 원리를 취하느냐에 따라 약간씩 달라질 수도 있으므로 음소 목록의 선정과 음소의 수도 다소간 차이가 있을 수 있다. 또한 시대적 차이, 방언적 차이, 연령적 차이, 계층적 차이에 의해서도 음운 체계 상의 변동이 다소간 있을 수 있다. 본 논문에서 인식 대상으로 선정한 음소 목록은 표 4.1과 같다. 음소의 분류 기준은 자음의 경우 조음 방식(manner of articulation)에 따라 파열음(stop 또는 plosive), 마찰음(fricative), 파찰음(affricative), 유음(liquid), 비음(nasals)으로 나뉘며, 모음의 경우 단모음(monophthong)과 복모음(diphthong)으로 분류하여 각 음소군에 대한 특징 지도를 구성하였다. 본 실험에서는 음소의 성질이 다른 종성과 초성을 구별하였으며 초성의 경우 파열음은 무성음만 고려하

표 4.1 한국어 음소군
Table 4.1 Korean phoneme classes

음소군 / 조음방식		종 류	갯수
파열음	평음	ㅂ, ㄷ, ㄱ	3
	경음	ㅃ, ㄸ, ㄲ	3
	격음	ㅍ, ㅌ, ㅋ	3
유, 비음		ㅁ, ㄴ, ㄹ(l/r), (ㅇ)	4
마찰음		ㅅ, ㅆ, ㅎ	3
파찰음		ㅈ, ㅉ, ㅊ	3
모음	단모음	아, 어, 오, 우, 으, 이, 에	7
	복모음	야, 여, 요, 유, 예, 와, 워, 의, 웨, 위	10
종성	내파음	ㄱ, ㄷ, ㅂ	3
	유, 비음	ㅁ, ㄴ, ㄹ, ㅇ	4
전체 인식대상 음소			43

였다. 모음은 단모음의 /에/, /애/는 통합하였고, 복모음의 /왜/, /웨/, /외/와 /에/, /애/도 동일 모음으로 취급하였다. 그 결과 인식 대상 음소의 수는 모두 43개로 정리되었다.

4.2 신경망 입력을 위한 전처리 및 특징 추출

음성 신호는 다음의 전처리 및 특징 추출 과정을 거쳐 신경망의 입력 벡터로 사용하였다.

1. 대역 통과 필터 : 70Hz~3.4KHz
2. A/D 변환 : 8KHz 샘플링(sampling), 16 비트 양자화(resolution)
3. Pre-emphasis : $H(z) = 1 - 0.95z^{-1}$
4. Hamming 창(분석) : $W(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1})$

   여기서 N = 100, 프레임(frame) 길이 = 10ms
5. 12차 LPC 계수
6. 12차 LPC cepstrum 계수 ($c_1 \sim c_{12}$)
7. -1~1 로 정규화(normalization)

음소 학습 및 테스트용 특징 추출 환경은 다음과 같이 작성하였고 각 음소의 지속 시간 길이와 유효 특징 추출 구간 정보를 이용하여[13] 수동으로 잘라내어(segmentation) 실험하였다. 하나의 동일한 음소가 전,후의 음소에 의하여 음성 파형으로서의 물리적 특성(스펙트럼, 지속 시간 등)이 다르게 실현되는 현상을 조음 결합(coarticulation)이라 하는데, 음소 중 특히 자음은 음소 환경에 따라 광범위하게 변화하는 사실을 감안하여 되도록 이러한 조음 결합의 영향이 작은 환경에서 단음절(C^iVC^f) 데이타 베이스를 구성하였고, 화자의 수는 1인으로 발음했다.

1. 초성 자음 (C^iV 음절, $C^f = 0$)을 위한 음절 구성
   초성 자음(C^i) 19개 * 단모음(V) 7개 * 5회 발성 = 665음절
2. 모음 (V, C^i, $C^f = 0$)
   (1) 단모음 7개 * 25회 발성 = 175음절
   (2) 복모음 10개 * 25회 발성 = 250음절
3. 종성 (VC^f 음절, $C^i = 0$)
   단모음(V) 7개 * 종성(C^f) 7개 * 5회 발성 = 245음절

따라서 음소 추출을 위한 음소 환경은 모두 1335음절이다.

4.3 SOFM 신경망 학습을 위한 파라미터 설정

1. 연결 강도의 초기화

Kohonen은 0.45~0.55 사이의 값으로 초기화 하였다[5]. 본 논문에서는 0~1사이로 초기화 하였다.

2. 경쟁 층의 노드 수

각 패턴 부류간의 경계선을 명확히 하기위해 충분한 갯수의 노드가 필요하다.

본 실험에서는 6*6(자음군), 8*8(모음군)의 정사각형 구조를 사용하였다.

3. 학습률 계수 $\alpha(t)$

$0 < \alpha < 1$을 만족하도록 초기값 $\alpha(0) = 0.3$으로 하였으며 시간이 지남에 따라 선형적으로 감소하도록 $\alpha(t) = \alpha(0)(1 - \frac{t}{T_{segment}})$를 적용하였으며 선형 시간 구간 $T_1 = 1000$으로 하였고, T의 증가율은 5, 학습률 감소율은 0.2로 주었다. 이렇게 함으로써 연결 강도의 진동(weight bouncing), 재정규화 문제를 최소화하고자 하였다.

4. 이웃(neighborhood)의 범위 $NE_c(t)$

이웃의 범위가 너무 작으면 위상학적 배열이 이루어지지 않을 수도 있으므로 $NE_c(0)$는 경쟁 층의 한 변의 길이의 1/2로 하였으며 선형 시간 구간 T_1에서는 0.5의 축소율로 학습시켰고 그 이후에는 NE_c를 1로 고정시켰다.

5. 반복 횟수(Iteration)

전체적으로 반복 횟수는 2000번으로 설정하였다. Kohonen은 전형적으로 10,000~100,000번의 반복을 권한다.

그림 4.1에 학습 파라미터 $\alpha(t)$와 $NE_c(t)$을 나타내었다.

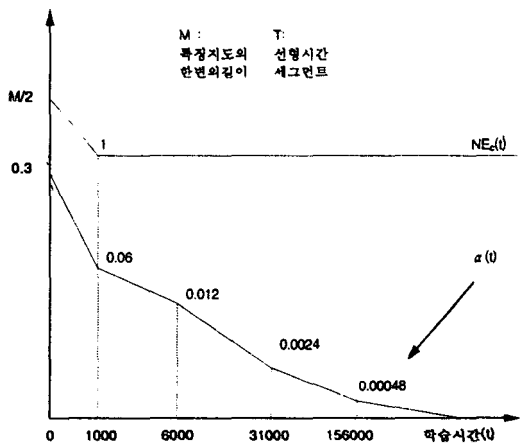


그림 4.1 학습 파라미터$\alpha(t)$와 $NE_c(t)$
Fig 4.1 Learning parameter $\alpha(t)$ and $NE_c(t)$

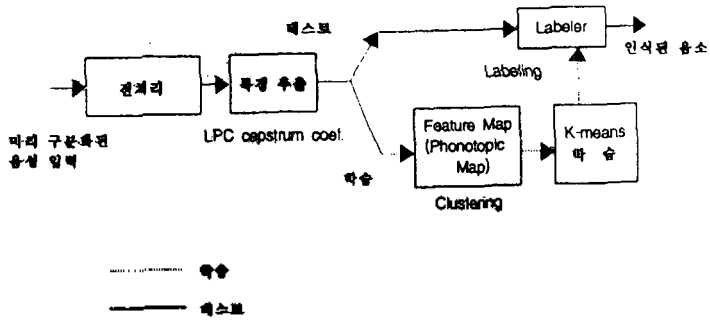


그림 4.2 제안된 음소 인식 시스템의 블럭도

Fig 4.2 Block diagram of the proposed phoneme recognition systen

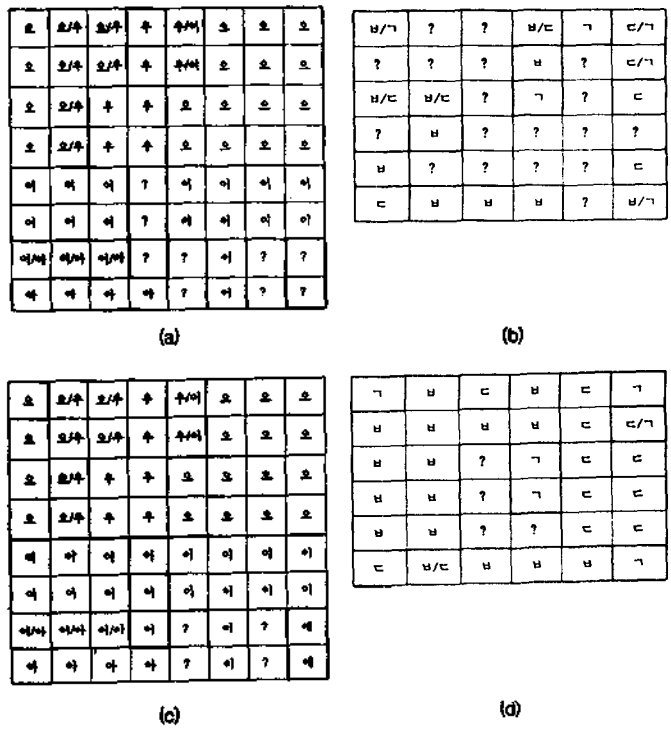


그림 4.3 K-means 학습 전,후의 음소 지도의 예

(a) K-means 학습 전의 단모음 지도

(b) K-means 학습 전의 /ㅂ, ㄷ, ㄱ/ 지도

(c) K-means 학습 후의 단모음 지도

(d) K-means 학습 후의 /ㅂ, ㄷ, ㄱ/ 지도

Fig 4.3 The phonotopic maps before and after K-means training

(a) The monophthong map and

(b) the /B, D, G/ map before K-means training

(c), (d) maps after K-means training

4.4 K-means 학습 결과

그림 4.2는 제안된 음소 인식 시스템의 블럭도를 나타낸다. 그림 4.3에 학습이 끝난 후에 경쟁 층 노드의 음소 지도(Phonotopic map)가 나타나 있다. 각 노드에 대응하는 음소 이름은 학습이 끝난 후에 연결 강도 값과 입력 벡터를 비교하여 수동으로 레이블링(labeling) 하였다. (a)는 단모음의 음소 지도며 (b)는 /ㅂ, ㄷ, ㄱ/의 음소 지도이다. (c)와 (d)는 K-means 학습에 의해 지도의 경계선이 분명해짐을 보여 준다. 이것은 학습이 불충분한 노드의 연결 강도를 재조정한 K-means 알고리즘의 유용성을 증명해 준다. 따라서 인식 성능의 향상과 빠른 수렴성을 가져오는 역할을 하게된다. 그림 4.3에서 ? 표시는 노드의 연결 강도가 최적화되지 않았음을 의미한다.

4.5 각 음소군 별 인식 결과 및 고찰

각 음소군에 대한 인식 실험은 화자 종속(speaker-dependent)으로 이루어졌으며 음소군내(intraclass)에서의 음소 상호간의 거리는 매우 가깝기 때문에 선택한 신경망의 패턴 분류 능력을 시험하기에 적절하다고 볼 수 있다. 음소군 별 인식 결과는 표 4.2와 같다.

인식 대상으로 한 전체 음소 토큰(token)의 수는 모두 1396개 였으며, 이 중 약 30%인 407개는 학습에 사용하였으며 학습에 참가하지 않은 나머지 70%인 989개를 테스트 대상으로 하였다. 학습 토큰의 경우 모음군에 대해서는 5개로 하였고, 자음의 경우 후 모음과의 조음 결합을 고려하여 14개(단모음 7개*2회 발성)를 학습 데이타로 선정하였다.

실험 결과 모음군에 대해서는 전체적으로 95.9%로 분류 능력이 우수했으며, 특히 단모음의 경우 인식이 잘 되었다. 복모음의 경우 /요/를 /여/로, /의/를 /위/로 오인식하는 문제가 생겼는데 이는 학습 패턴 선정시의 부주의로 여겨진다. 자음군에 대해서는 전체적으로 82.6%의 낮은 인식률을 보였으며 파열음의 경우 후 모음과의 조음 결합이 심해서 전반적으로 각 음소의 유효 특징 구간(steady-state segment)을 잡아내기가 어려웠다. 짧게는 1프레임 미만인 것부터 길게는 3프레임 길이까지 분포 영역이 다양했다. 실험시 선택한 파열음의 특징 구간은 시단에서 2프레임으로 정하였다. 실험 결과 82.7%의 인식률을 얻었다. 유음 및 비음의 경우 특징량으로 선택한 켑스트럼(cepstrum)의 문제점이 보였는데, /ㅁ/, /ㄴ/같은 비강과 구강의 공진 특성을 갖는 비음의 경우 구강(vocal tract)만을 모델링하는 특징 계수 자체의 문제점으로 인해서 /ㄴ/과 /ㄹ(l)/의 경우 조음점(articulation point)이 유사한 관계로 혼동이 심했다. 한편, /ㄹ(r)/의 경우 분류 능력이 우수했는데(95.2%), 이는 혀굴

표 4.2 각 음소군에 대한 음소 인식 결과
(a) 모음 전체 (b) 파열음
(c) 유, 비음/마찰음/파찰음 (d) 종성/전체

Table 4.2 Korean phoneme recognition results for each class
(a) Vowels (b) Plosives
(c) Liquids, nasals/fricatives/affricatives
(d) Final consonants and the whole

음 소	토 큰 수		에러수	인 식 률	
	학 습	테스트			
단 모 음					
아	5	20	0	100%	100%
어	5	20	0	100%	
오	5	20	0	100%	
우	5	20	0	100%	
으	5	20	0	100%	
이	5	20	0	100%	
에	5	20	0	100%	
복 모 음					
야	5	20	1	95%	93%
여	5	20	1	95%	
요	5	20	5	75%	
유	5	20	0	100%	
예	5	20	0	100%	
와	5	20	2	90%	
워	5	20	1	95%	
의	5	20	4	80%	
웨	5	20	0	100%	
위	5	20	0	100%	
모 음 전 체				95.9%	

(a)

음 소	토 큰 수		에러수	인 식 률	
	학 습	테스트			
파 열 음					
ㅂ	14	21	3	85.7%	82.7%
ㄷ	14	21	4	81%	
ㄱ	14	21	3	85.7%	
ㅍ	14	56	8	85.7%	
ㅌ	14	56	14	75%	
ㅋ	14	56	9	83.9%	
ㅃ	14	21	2	90.5%	
ㄸ	14	21	4	81%	
ㄲ	14	21	4	81%	

(b)

유, 비음, 마찰음, 파찰음					
ㅁ	14	21	5	76.2%	79.8%
ㄴ	14	21	6	71.4%	
ㄹ(l/r)	l/14	21	5	76.2%	
	r/14	21	1	95.2%	
ㅅ	14	21	4	81%	82.5%
ㅆ	14	21	5	76.2%	
ㅎ	14	21	2	90.5%	
ㅈ	14	21	2	90.5%	95.2%
ㅉ	14	21	1	95.2%	
ㅊ	14	21	0	100%	

(c)

종		성			
ㄱ	14	21	4	81%	78.6%
ㄷ	14	21	5	76.2%	
ㅂ	14	21	6	71.4%	
ㄴ	14	21	3	85.7%	
ㄹ	14	21	6	71.4%	
ㅁ	14	21	3	85.7%	
ㅇ	14	19	4	71.4%	

전 체	토 큰 수		에러수	인식률
	학 습	테스트		
	407	989	127	87.2%

(d)

림 소리 특징을 잘 포착한 것으로 보여진다. 마찰음군의 경우 특히 /ㅅ/과 /ㅆ/의 분류 능력이 떨어졌는데, 두 음소의 지속 시간 길이가 대부분 겹치고 특징 계수의 패턴이 유사해 혼동이 심했다. 그러나 /ㅎ/의 경우 분류 능력이 우수했는데(90.5%), 기음의 특징 패턴이 앞의 두 음소와는 현저히 다름에서 기인한 것으로 생각된다. 파찰음군의 경우는 특징 계수 패턴을 관찰한 결과 지속 시간 길이 정보는 5프레임~11프레임의 7프레임 길이에서, 특징 계수 추출은 C_1~C_5의 정보가 인식에 유효함을 알 수 있었다. 파찰음의 경우 95.2%의 인식률을 보여 분석 결과가 타당함을 확인하였다. 종성의 경우 비교적 지속 시간 길이가 긴 /ㄴ, ㄹ, ㅁ, ㅇ/의 경우 유음 및 비음의 경우처럼 /ㄹ, ㅇ/과 /ㄴ/의 혼동이 심했으며, 내파음(implosive)인 /ㄱ, ㄷ, ㅂ/의 경우 지속 시간 길이도 비교적 짧고, 앞 모음에서 폐쇄되는 과정으로의 전이 구간(transient region)이 길어 유효 특징 구간 설정이 힘들었다. 실험시 설정한 특징 구간을 종단에서 3프레임으로 하였는데, 전이 구간을 포함시켜 실험한 결과 분류 능력이 현저히 떨어짐을 확인하였다. 실험 결과 78.6%의 인식률을 얻었다.

각 음소군별 전체에 대한 인식 결과 평균 시스템의 성능은 87.2%의 인식률을 보였으며 다음과 같은 결론을 얻었다.

V. 결 론

첫째, 본 네트웍은 다층 인식자(Multilayer Perceptron)가 패턴 공간을 무리하게 구분지으려는데서 문제를 일으키는것(overspecialization)과는 달리 구분적 선형(piecewise linear)식별 함수를 형성하면서도 분류 능력이 우수하다는 점,

둘째, 특징 추출 단계에서 얻어진 특징량에는 그것이 인식에 유효한 특징이라는 적극적 의미가 부여되지는 않으므로 앞으로는 사용하는 특징량에 대한 세밀한 분석을 통해 우리말 인식에 필요한 특징을 선택할 필요(feature selection)가 있다는 점(예, 파찰음군 인식 실험 결과)

셋째, 유사 음소군내(intraclass)에서만 인식 실험을 할 때에는 각 음소의 유효 특징 구간(steady-state segment) 정보가 중요하지만 다른 음소군간(interclass) 테스트시에는 자음의 경우 조음 결합에 따른 전이 구간(transient region)의 정보가 식별에 중요한 요소가 되므로 우리말 음소의 특징 패턴을 명확히 규명하는 작업이 선행되어야 할 필요가 있다는 점,

넷째, 현실적으로 각 음소군간의 겹침(overlap)에 따른 인식 불능(reject) 영역이 존재하므로 신호 레벨에서 100%의 인식률을 기대하기는 어렵다는 점,

다섯째, 모든 신경망 접근법이 그렇듯이 네트웍 학습시 사용되는 학습 패턴설정이 분류 성능을 좌우하므로 음소 학습 패턴의 경우 구분화(segmentation) 단계에서 주의가 필요하다는 점이다. 본 논문에서 구현한 신경망은 자기 연상(autoassociation)을 수행하는 대표적인 네트웍이므로 그 중요성이 크다고 하겠다.

음성 인식의 최종 목표인 높은 인식 성능, 대 용량 어휘, 화자 독립, 연속 음 성 인식을 수행하는 첫 번째 단계는 인식 성능의 개선이다. 특히 음소를 기반으로 하는 인식 시스템은 음소 분류 능력이 좋아야 한다. 물론 음소 인식 단 계에서 100%의 인식률을 기대할 필요는 없는 것으로 보인다[16]. 왜냐하면 인간은 개개의 음소 인식률이 완전하지 않더라도 문맥

등의 상위 차원 정보를 이용하여 문장을 '이해'하고 있기 때문이다.

본 연구에서는 우리말 인식에 대한 기초 연구로서 먼저 인식의 변별 단위 로서 음소를 정의, 분류하고 이를 각각의 음소군으로 분리한 뒤 SOFM 신경 망으로 인식하는 실험을 하였다. 각 음소군 별로 전체 인식률은 비교적 높게 나타났으며 테스트 데이타에 대한 화자 종속 실험 결과 87.2%의 비교적 높은 인식률을 얻어 음소 인식의 가능성을 보였다. 그러나 '귀'의 능력만으로는 인식, 이해가 불가능하므로 앞으로는 '두뇌'의 능력, 즉 상위 차원에서 하부의 오류를 흡수, 교정할 수 있는 우리말 음운 현상의 체계를 규칙(rule)화한 규칙 베이스(Rule-Base)와의 결합을 통해 해결해야 할 것이다.

참 고 문 헌

1. R. Schalkoff, *Pattern Recognition*, John Wiley & Sons, 1992.
2. D. P. Morgan, C. L. Scofield, *Neural Networks and Speech Processing*, Kluwer Academic, 1991.
3. R. P. Lippmann, B. Gold, "Neural-net classifiers useful for speech recognition", *Proc. of the 1st IEEE ICNN*, San Diego, Vol. IV, pp. 417-425, 1987.
4. H. P. Siemon, "Selection of Optimal Parameters for Kohonen Self-Organizing Feature Maps", *Artificial Neural Networks*, Vol. 2, pp. 1573-1577, 1992.
5. T. Kohonen, *Self-Organizing and Associative Memory*, 2/e, Springer-Verlag, Berlin 1987.
6. P. Brauer, "Infrastructure in Kohonen Maps", *ICASSP*, Vol. 1, pp. 647-650, 1989.
7. I. Aleksander(Ed.), *Neural Computing Architecture*, Chap.3, Chap.4, The MIT Press, 1989.
8. P. K. Simpson, *Artificial Neural Systems*, Pergamon Press, 1990.
9. J. A. Freeman, et al., *Neural Networks*, Chap. 2, Chap. 6, Chap. 7, Addison-Wesley, 1991.
10. B. Kosko, *Neural Networks for Signal Processing*, Chap. 1, Prentice-Hall, 1992.
11. Z. Huang, A. Kuh, "A Combined Self-organizing Feature Map and Multilayer Perceptron for Isolated Word Recognition", *IEEE Trans. on Signal Processing*, Vol. 40, No. 11, Nov. 1992.
12. 성 백인, 김 현권, *언어학 개론*, 한국 방송통신대학, 1992.
13. 김 범국, 정 현열, "한국어 단음절에 포함된 음소 인식에 관한 연구", *한국 음향학회, 제 9 회 음성 통신 및 신호 처리 워샵 논문집*, 제 SCAS-9 권 1호, 1992.
14. R. Beale, T. Jakson, *Neural Computing: An Introduction*, Adam Hilger
15. T. Kohonen, "The Neural Phonetic Typewriter", *Computer Magazine*, pp. 11-22, Mar. 1988.
16. G. A. Carpenter, S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*, Chap. 5, The MIT Press, 1991.
17. 권 영욱, 정 현열, "MDS법을 이용한 한국어 단모음의 분석", 한국 음향학회, 국제 음향 학술 발표회 논문집, 1990, 11.
18. 구 명완, 은 종관, "LVQ를 이용한 화자 적응에 관한 연구", 한국 음향학회, 국제 음향 학술 발표회 논문집, 1990, 11.
19. 이 종락, "반음소 : 새로운 음성 합성 및 인식 단위", 한국 음향 학회, 제10회 음성통신 및 신호처리 워샵 논문집, 제 SCAS-10권 1호, 1993.

▲전 용 구(Yong-Koo Jeon)

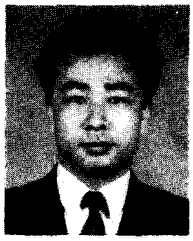
1992년 2월 : 광운대학교 전자계산기공학과 졸업(공학사)

1994년 2월 : 광운대학교 대학원 전자계산기공학과 졸업(공학석사)

현재 : 세일 정보통신사 시스템개발 2부

※주관심분야 : Neural network, Speech signal processing 등

▲양 진 우(Jin-Woo Yang : 정회원, 종신회원)



1959년 9월 30일생

1982년 2월 : 원광대학교 전자공학과 졸업(공학사)

1985년 2월 : 광운대학교 대학원 전자공학과 졸업(공학석사)

1994년 8월~현재 : 광운대학교 대학원 전자계산기공학과(박사과정수료)

1993년 7월~1994년 3월 : 광운대학교 전자계산기공학과 전임조교
※주관심분야 : Voice dialing, Real-time processing, Neural network, Speech recognition & synthesis, Speech signal processing 등

▲김 순 협(Soon-Hyob Kim)
현재 : 광운대학교 컴퓨터공학과 교수
한국음향학회지 13권 6호 참조.