1993년도 한국음향학회 학술논문발표회 논문집(제 12권 1(s)호)

SOFM 신경회로망을 이용한 한국어 음소 인식

전 용구°, 양 진우, 김 순협
광운대학교 컴퓨터 공학과

Korean Phoneme Recognition Using Self-Organizing Feature Map

Yong-Koo Jeon°, Jin-Woo Yang, Soon-Hyob Kim
Dept. of Computer Engineering, Kwang Woon Univ.

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 clusterer로 사용하였다. SOFM 신경망은 신호 공간에 대해서 최적의 국소(局所) 해부적 사상에 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 평가하기위해 먼저, 우리말 음소들을 모음, 파열음, 마찰음, 파찰음, 유음 및 비음, 종성의 6개 음소군으로 분류하고 각 음소군에 대한 특징 지도를 구성하여 labeler의 기능을 수행하게 하였다. 화자 종속 인식 실험 결과 87.2%의 인식률을 보였으며 제안한 학습법의 빠른 수렴성과 인식률 향상을 확인하였다.

I. 서 론

음성 인식의 방법은 주로 패턴 정합(pattern matching) 법에 근거를 두고 있으며, 에널로그, 디지틀 VLSI 제조 기술의 발전으로 병렬 처리가 가능해짐으로써 인간의 두뇌를 모방한 신경 회로망(Neural Network)을 이용한 음성 인식 기술이 대두되고 있다. 한편 음성 인식의 최종 목표는 (1) 불특정 화자(speaker independent) (2) 연속 음성 (3) 대 어휘의 실현으로 볼 수 있는데 대 어휘 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(dipbone), 음절(syllable) 등의 단어 하부 단위(sub-word unit)로 모델링하게 된다. 특히 형태소의 결합 유형에서 볼 때 교착어(agglutinative language)에 속하는 한국어는 그 어형 성법에 있어서 굴절, 파생, 합성 세 가지를 모두 널리 사용하므로 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 clusterer로 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-음소(quasi-phoneme)로 변환시키는 labeler의 역할을 수행하는 것이다.

II. 신경 회로망을 이용한 음성 인식

우리가 인식 대상으로 하는 정보를 표현하며 전달하기 위해 인간이 의도적으로 발생시키는 패턴의 대표적인 예는 음성이다. 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 확률론적(statistical/decision theoretic) 또는 결정 이론적 방법과 구조 해석적(syntatic) 방법의 두 가지가 있어왔다. 최근에 각광받는 신경망(Neural Network) 기술은 그 세 번째 접근 방법으로서 음성 인식에 있어서 기존의 패턴 생성 메카니즘(mechanism)을 모델링하는 것을 피하고 대신에 입/출력으로부터 또는 'black box'의 관점에서 문제를 다루는 것이다. 즉 인간의 두뇌는 입/출력 특징을 정량화하는 자세한 알고리즘 세트가 없이도 지능적인 동작(패턴 인식과 분류를 포함하는)을 관찰하고 흉내낼 수 있다는 사실로 부터 좋은 black box 모델이 된다고 할 수 있다. 이러한 신경망적 접근법의 유용성은 패턴 분류에 필요한 다양한 식별 함수(discriminating function)의 생성과 모든 패턴을 미리 조사할 수는 없으므로 학습 패턴(training pattern)으로 부터 식별 함수를 결정하기 위한 학습(learning)이 자유롭다는 점을 들 수 있다.

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

Ⅲ. K-means 알고리즘이 결합된 SOFM 신경 회로망

인간 두뇌의 정보 처리 과정을 살펴보면 특별한 교사(teacher) 신호가 없이도 외계의 정보 신호에 따라 그것을 뇌 속에 표현하는 자율 학습(unsupervised learning)의 메카니즘을 가지고 있음을 알 수 있다. 예를 들면 대뇌 피질에서는 외계 신호 공간의 정보 구조를 2차원인 신경장(fields, layer, slab)으로 사상(mapping)하여 표현하는 자기 조직화(self-organization) 과정의 결과로 정보를 국재(局在)적(local)으로 표현하는 것을 볼 수 있다. 이러한 자기 조직화 과정은 레이블링(labeling)되지 않은 데이타를 부분 집합으로 분할(partition)하는 nonparametric 접근법인 기존의 클러스터링(clustering)기법과 동일하며, 대표적인 알고리즘으로 반복적(iterative) 접근법인 K-means 알고리즘을 들 수 있다. 여기서 K란 단순히 부류의 갯수를 가리킨다. 반복적 접근법의 경우 클러스터링 평가 함수(criterian function) 또는 성능 지수(performance index)의 최소화/최대화에 기초하여 스스로 반복적 절차를 통해 주어진 데이타에 내재하는 자연적인 그룹들(natural groups/clusters)을 발견하게 된다. 그러나 K-means 알고리즘은 다음 네 가지의 제약을 지니고 있다.

(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 'Cluster validity'의 연구 분야로서 흥미있는 주제가 되고 있다. 본 논문에서는 이러한 네 가지 제약성을 SOFM 신경 회로망을 통해 해결하고자 하였다.

SOFM 신경망은 입력 정보를 2차원의 신경장에 국재화(localization)시켜 응답하게하는 메카니즘인 경쟁 학습(competitive learning)을 통해 입력 패턴내에 존재하는 더욱 구조(structure)를 발견하는 특징 검출기(feature detector)의 역할을 하게 된다. 본 논문에서 사용한 SOFM 신경 회로망의 구조는 그림 1과 같으며 2차원의 평면 위상(planar topology)을 형성하는 노드의 연결 구조를 갖는다. 물론 3차원과 그 이상의 차원을 갖는 위상도 가능하지만 P. Brauer 등의 연구에 의하면 2차원 위상보다 성능이 떨어지는 것으로 보고되었다.

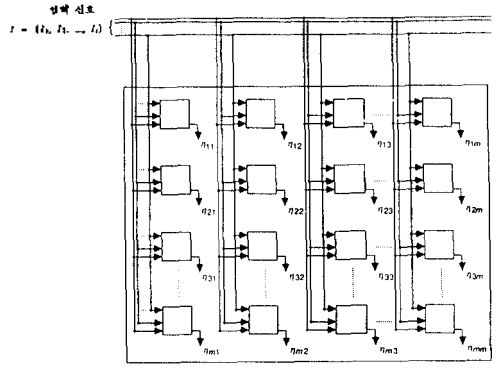

그림 1. SOFM 신경 회로망의 구조

이것은 d차원의 특징 공간(feature space)이 자기 조직화 특성인 차원 축소(dimensionality reduction) 결과 2차원의 특징 지도상에 사상됨을 의미한다. 축소된 차원 공간에서 노드들간의 위상학적 거리는 비유사도(dissimilarity)에 비례한다. 본 논문에서 유사도 계산을 위해 유클리드 노름(Euclidean Norm)을 사용하였는데 유크리드 거리 계산을 위해 취하는 제곱근을 벗기면 $d^2(I, W)$는 입력 벡터와 연결 강도 벡터의 상관(correlation)을 반영하며 정규화된 정합(normalized matching)을 나타낸다. 이것은 입력 벡터와 연결 강도 벡터간의 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 MSE)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 감강법(Gradient-descent Rule)임을 기억하면 연결 강도 벡터 W는 그 자신과 입력 벡터 I사이의 오차를 가장 급한 기울기 방향으로 줄여 가는 것을 알 수 있다. 따라서 유클리드 거리식은 네트웍의 연결 강도 수정을 위한 오차 함수(error function)로 해석되며, 학습은 연결 강도 벡터의 최적화(optimization)문제로 귀결된다.

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

식 1. 학습을 위한 오차 함수

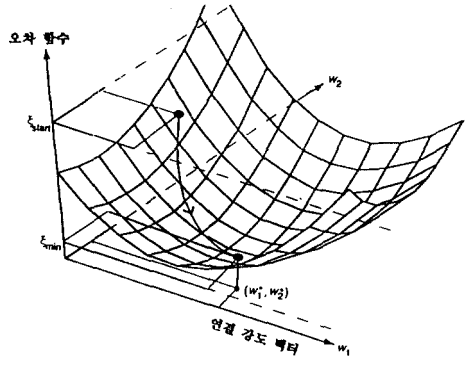

그림 2. Gradient-descent rule에 의한 연결 강도 조정

위에서 언급한 바와같이 SOFM 신경망의 학습이 오차의 제곱 평균(MSE)을 최소화시키는 과정으로 해석될 수 있지만 (그림 2) 실제로 학습 과정을 중지시키는 판단 기준은 학습률 계수 또는 반복 횟수가 되며 물론 이것은 네트웍이 충분히 학습되었음을 보장해주지 않는다. 예를 들면 학습 패턴 세트가 적절히 부류를 나타내지 못하는 경우 출력 층의 연결 강도 벡터는 정확한 중심점(mean, centroid, reference, codeword, template)을 찾아내지 못하게 되므로 결국 충분한 학습이 되지 않았음을 의미한다. 이러한 상황이 그림 3에 나타나 있다. 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 네트웍의 연결 강도 벡터를 미세 조정(fine-tuning)할 필요가 있다.

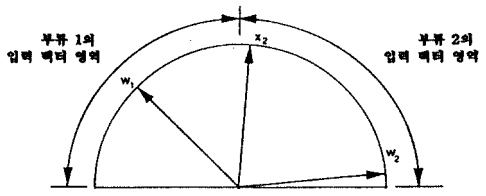

그림 3. 국부 최소점 문제에서 기인하는
오분류의 예

본 논문에서는 K-means 클러스터링에서 도입했던 오차의 제곱 합(Sum of Squared Error) 평가 함수를 최소화하도록 특징 지도의 연결 강도 벡터를 미세 조정함으로써 국부 최소점(local minima)문제를 해결하고자 하였다. 즉 특징 지도의 오차 제곱의 합 또는 전체 '분산(variance)'을 다음과 같이 정의 하고

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

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 패턴 세트 S_p의 중심점(centroid)으로 바꾼다.

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

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.

본 논문에서 구현한 음소 인식 시스템은 그림 4와 같다.

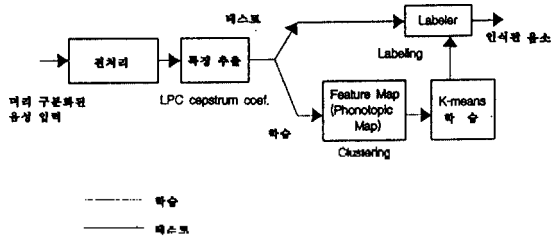

그림 4. 제안된 음소 인식 시스템의 블럭도

Ⅳ. 시뮬레이션 결과 및 고찰

본 논문에서 인식 대상으로 선정한 음소는 조음 방식(manner of articulation)에 따라 모음, 파열음, 마찰음, 파찰음, 유음, 비음, 중성의 6개 음소군으로 분류하고 각 음소군에 대한 음소 지도(phonotopic maps)를 구성하였다. 본 실험에서는 음소의 성질이 다른 중성과 초성을 구별하였으며 초성의 경우 파열음은 무성음만 고려하였다. 모음은 단모음와 /에//애/는 통합하였고, 복모음의 /왜//웨//외/와 /예//에/도 동일 모음으로 취급하였다. 그 결과 인식 대상 음소의 수는 모두 43개로 정리되었다. 음소 학습 및 테스트용 특징 추출 환경은 조음 결합(coarticulation)현상을 최소화하도록 1335개의 단음절(C^1VC^f) 데이타 베이스를 구성하여 각 음소의 지속 시간 길이와 유효 특징 추출 구간 정보를 이용하여 수동으로 잘라내었다(segmentation). 특징 계수로는 12차의 LPC cepstrum 계수(C_1-C_{12})를 구하여 신경망에 입력하기 위해 -1~1로 정규화(normalization)하였다. 그림 5는 K-means 학습결과 미세조정된 음소 지도를 보여주며 K-means 학습에 의한 네트워의 빠른 수렴성을 확인하였다.

각 음소군에 대한 인식 실험은 화자 종속(speaker dependent)으로 이루어졌으며 음소군내(intra-class)에서의 음소 상호간의 거리가 매우 가깝기 때문에 선택한 신경망의 패턴 분류 능력을 시험하기에 적절하다고 볼 수 있다. 각 음소군에 대한 인식 실험 결과 모음 전체에 대해서 95.9%, 자음 전체에 대해서 82.6%, 그리고 음소군 전체에 대해서는 87.2%의 비교적 높은 인식률을 얻을 수 있었으며 다음과 같은 결론을 얻었다.

*** /ㅂ,ㄷ,ㄱ/ Map ***
(K-means Training 전)

ㅂ/ㄱ	?	?	ㅂ/ㄷ	ㄱ	ㄷ/ㄱ
?	?	?	ㅂ	?	ㄷ/ㄱ
ㅂ/ㄷ	ㅂ/ㄷ	?	ㄱ	?	ㄷ
?	ㅂ	?	?	?	?
ㅂ	?	?	?	?	ㄷ
ㄷ	ㅂ	ㅂ	ㅂ	?	ㅂ/ㄱ

*** /ㅂ,ㄷ,ㄱ/ Map ***
(K-means Training 후)

ㄱ	ㅂ	ㄷ	ㅂ	ㄷ	ㄱ
ㅂ	ㅂ	ㅂ	ㅂ	ㄷ	ㄷ/ㄱ
ㅂ	ㅂ	?	ㄱ	ㄷ	ㄷ
ㅂ	ㅂ	?	ㄱ	ㄷ	ㄷ
ㅂ	ㅂ	?	?	ㄷ	ㄷ
ㄷ	ㅂ/ㄷ	ㅂ	ㅂ	ㅂ	ㄱ

그림 5. K-means 학습 전,후의 음소 지도의 예

Ⅴ. 결 론

첫째, 본 네트워은 다층 인식자(Multilayer Perceptron)가 패턴 공간을 무리하게 구분지으려는데서 문제(overspecialization)를 일으키는 것과는 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 특징량(예를 들면, cepstrum coef.)에 대한 세밀한 분석을 통해 우리말 인식에 필요한 특징을 선택(**feature selection**)할 필요가 있다는 점,

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 전이 구간(**transient region**)의 정보가 식별에 중요한 요소가 되므로 우리말 음소에 대한 특징 패턴을 명확히 규명할 필요가 있다는 점,

넷째, 현실적으로 각 음소군이 특징 공간상에서 서로 겹치므로(**overlap**) 인식 불능 영역(reject)이 존재하므로 신호 레벨에서 100%의 인식률을 기대하기 어렵다는 점,

다섯째, 모든 신경망 접근법이 그렇듯이 네트웍 학습시 사용되는 학습 패턴 설정이 정확할 수록 분류 성능이 좋아지므로 음소 학습 패턴의 경우 구분화(segmentation) 단계에서 주의가 필요하다는 점이다. 본 논문에서 구현한 신경망은 자기연상(**autoassociation**)을 수행하는 대표적인 네트웍이므로 그 중요성이 크다고 하겠다.

본 연구에서는 우리말 인식에 대한 기초 연구로서 먼저 인식의 변별 단위로서 음소를 정의, 분류하고 이를 각각의 음소군으로 분리한 뒤 SOFM 신경망으로 인식하는 실험을 하였다. 각 음소군별로 전체 인식률은 비교적 높게 나타났으며 테스트 데이타에 대한 화자 종속 인식 실험 결과 87.2%의 인식률을 얻어 음소 인식의 가능성을 보였다. 앞으로는 '귀'의 능력만이 아닌 '두뇌'의 능력, 즉 상위 차원에서 하부의 오류를 흡수, 교정할 수 있는 우리말 음운 현상의 체계를 규칙(Rule)화한 지식베이스(Knowledge-base)와의 결합이 중요하리라 생각된다.

표 1. 각 음소군에 대한 음소 인식 결과
(a) 모음 전체 (b) 파열음 (c) 유, 비음/마찰음/파찰음
(d) 종성/ 전체

음 소	토큰수 학습	토큰수 테스트	에러수	인식률	
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	토큰수 학습	토큰수 테스트	에러수	인식률	
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	토큰수 학습	토큰수 테스트	에러수	인식률
	407	989	127	87.2%

(d)

참고 문헌

1. R. Schalkoff, *Pattern Recognition*, John Wiley & Sons, 1992.
2. J.R. Deller Jr., et al., *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.
3. B.W. Wah, "Special Issue on Artificial Neural Networks Guest Editor's Introduction", *IEEE Trans. on Computers*, Vol.40, No.12, Dec. 1991.
4. H. SPATH, *Cluster Analysis Algorithms*, Ellis Horwood, Limited, 1980.
5. C. von der Malsberg, "Self-organizing of orientation sensitive cells in the striate cortex", *Kybernetik*, 14, pp.85-100, 1973.
6. J. Kangas, T. Kohonen, et al., "Variants of Self-Organizing Maps", *IJCNN*, Vol.2, pp.517-522, 1989.
7. P. Brauer, "Infrastructure in Kohonen Maps", *ICASSP*, Vol.1, pp.647-650, 1989.

8. E. Saund, "Dimensionality-Reduction Using Connectionist Networks", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.2, No.3, March 1989.

9. T. Kohonen, "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, 1982.

10. 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.

11. A. Papoulis, *Probability, Random Variables, and Stochastic Process*, 3/e, McGraw-Hill, 1991.

12. P. Antognetti, V. Milutinovic (Ed.), *Neural Networks - Concepts, Applications, and Implementations*, Vol. 1, Chap.3, Chap4, Prentice-Hall Advanced Reference Series, 1991.

13. E. McDermott, S. Katagiri, "LVQ-based Shift-tolerant Phoneme Recognition", *IEEE Trans. on Signal Processing*, Vol. 39, No. 6, June 1991.

14. G. A. Carpenter, S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*, Chap.5, The MIT Press, 1991.