공학 연구논문집 제24권 제1호 pp. 109~120, 울산대학교
Journal of Engineering Research Vol. 24, No. 1, pp. 109~120, 1993.
University of Ulsan.

동적 신경회로망을 이용한 음성인식에 관한 연구

이태호
전자공학과

〈요 약〉

재귀형 인공신경망을 사용하는 음성인식 계통을 구성하고 동작을 분석하였다. 신경망으로는 부분결합재귀(PCR)망과 부분결합 Elman(PCE)망의 두가지 변형이 시험되었으며, 시간적 변화를 가지는 입력 시이퀀스에 대해서 어떻게 동작하는가를 중점적으로 고찰하였다. 음성 시료는 한국어 자음 "ㄱ"과 "ㄷ", "ㅂ"의 세가지를 대상으로 하고, 이들이 모음-자음-모음의 형태로 나타나는 여러 경우를 택하여 시료로 하였다. 교육된 시료에 대해서는 두 신경망 모두 100% 인식률을 보였으나 교육되지 않은 시험용 시료에 대해서는 PCR망은 62.2%, 그리고 PCE망은 73.3%의 인식률을 나타내었으며 사용된 시료들의 대단히 과도적인 성격이 인식률을 낮추는 주요 원인이라고 판단되었다. 실험결과의 분석으로부터 개선된 시스템을 위한 방향이 제시되었다.

A Study on the Speech Recignition Using Dynamic Neural Network

Lee Tai-ho
Deptartment of Electronic Engineering

〈Abstract〉

The phoneme recognition using the recurrent type neural networks has been studied. Two variations of network have been tested, that is: partially connected recurrent(PCR) network and partially connected Elman(PCE) network. The interest was focused on the behaviour of networks in response to the time varying

* 본 연구는 1992년도 울산공업학원 이사장 연구비의 지원으로 수행된 것임.

input sequences. The speech samples used for the recognition task were Korean phonemes "G", "D" and "B" taken in the varying context of VCV structure. Both of the network recognized all the training samples. For untrained test samples, the recognition rates of 62.2% and 73.3% were achieved by PCR and PCE network, respectively. Rather low rate of recognition is considered mainly due to highly transient nature of speech samples used. A discussion toward the improvement of network performance is made based on the analysis of experimental results.

I. 서 론

지난 이십여년간의 대량의 연구결과에도 불구하고 음성인식은 '아직 해결되지 못한 문제'로 남아 있음을 지적하고, Waibel[1]은 음성인식의 문제영역에 따른 난이도를 다음과 같이 제시하고 있다.

- 고립단어, 연결단어 및 연속어
- 어휘의 규모
- 화자독립과 종속
- 음향학적 애매성과 혼란정도
- 주변잡음

이들 각 영역에서 가장 낮은 수준에 대해서는 긍정적 성과가 이미 20년 전에 나타나 있었으나, 그 후의 발전이 지연되고 있음을 보면 위의 각 영역이 제기하는 문제는 양적이 아니라 질적인 것이어서 어떤 도약을 위한 혁신적 수단을 기대하고 있음을 알 수 있다. 실제의 시스템이 직면하는 문제의 본질을 정리해 보면 다음 세 가지로 요약된다. 첫째로 최전단부(front end)에서 당면하는 음향학적 정보의 불완전성 문제가 있다. 음운학적 기호, 즉 음소 등은 실제 발성에서는 여러가지로 변형되어, 길이가 변화하고 경계가 불확실하며 심지어 탈락되는 등 극히 왜곡된 음향정보의 형태로 제공된다. 이 신호로부터 어떻게 필요한 정보를 끌어내느냐가 문제인 것이다. 둘째로 반대쪽 측면에서 음성인식을 위한 사람의 지적 과정에 대한 이해가 부족할 뿐 아니라 이를 기계화할 기술도 없는 점이다. 즉 사람의 음성인식 과정에서는 자신의 지적 능력이 총체적으로 작용하여 빠진 소리나 틀린 말도 채워 넣을수 있는 일종의 생산적 과정이 형성되므로 이의 일부라도 기계화하는 것은 두뇌의 기능을 모형화하는 수준의 작업이 되는 것이다. 셋째로는 기계의 처리속도의 문제이며, 이 세가지 문제의 해결이 음성인식을 개선하는 요체가 된다. 70년대 이래 동적계획법(DP)과 HMM(hidden Markov model)기법이 도입되어 크게 공헌하여 왔는데[2][3] 이들은 앞의 첫째와 둘째 문제 영역의 상호작용을 통하여 해결에 접근하고 있는 것이다. 특히 HMM은 논리적 배경이 정연하고, 여러 계층으로의 확대 및 변형의 가능성이 다양하여 현재 가장 효과적인 방법으로 되어 있다.

80년대 이래 재조명을 받기 시작한 인공신경망(또는 연결주의 망)은 새로운 가능성으로 받아들여지고 있다.[4] 인공신경망의 동작은 본질적으로 분류기적 기능에 기초하고 있으며[5] 여러 계층에서 분류기능을 필요로 하는 음성인식 분야에서 관심의 대상이 될 수 밖에 없었다. 인공신경망의 매력 중 하나는 '학습'의 과정을 통하여 인식 대상의 내재적 구조를 스스로 파악하게 한다는 것이다. 네개의 층으로 구성된 다층 퍼셉트론과 오류역전파 알고리즘을 결합하여 자유로운 식별곡면을 구성할 수 있음이 밝혀져 있다. 80년대말로부터 현재까지 엄청나게 많

은 보고가 인공신경망에 관하여 발표되어 왔으나 실제로 음성인식에 나타난 공헌은 극히 미약하다. 그 이유로서는 우선 인공신경망의 내부적 동특성에 대한 이론적 해석이 거의 되어있지 않은 것을 들 수 있고 둘째로 대부분의 모형이 음성신호의 동적 성질을 수용하기에 적합하지 않은 정적 패턴 인식기의 범주에 속하고 있기 때문이다.[6][7] 이에 따라 본 연구에서는 부분결합 회귀신경망 계통[8]의 동적 신경망을 사용하여 음성인식을 시도하고, 현상을 분석하여 발전의 방향을 제시하고자 한다.

II. 시스템 구조

1. 망 구조

본 연구에서 목표한 것은 일반적 상황에서 음소를 인식하는 모형을 세워보고자 하는 것이었다. 단어의 발성에 포함되어 있는 음소들은 항상 똑같지 않을 뿐 아니라 발성도중에서도 변화가 존재하므로 큰 특징을 포착하는 데에는 동적신경망이 적합할 것이다. 동적 특성을 표현하기 위한 신경망의 두 유형으로서는 지연적 성격을 가진 것과 회귀적 성격의 것이 있는데, 지연적인 것은 망 자체로는 동적이라고 할 수 없으므로, 본 연구에서는 그림 1과 같은 회귀형을 택하였다. 이 모형은 기본적으로는 Elman망[9]의 형태이지만 입력층과 문맥층에서 은닉층 사이의 연결은 통상의 완전결합이 아닌 무작위 부분결합을 이루도록 하여 특성의 개선을 꾀하였다. 이 망을 PCE(partially connected Elman)망이라고 하기로 한다. 이와 같은 변형의 동기는 그림 2에 보인 부분결합 회귀신경망(PCRNN, 이하 PCR망)[8]의 연구 결과에 근거하였다.

PCR망의 기본적 특징은 삼중의 회귀구조와 부분결합을 말할 수 있다. 그 결과 시계열 정보의 처리기능이 Elman이나 Jordan 망에 비하여 우수한 것으로 보고되었으며 계산시간을 감축시키고 잡음에 대한 감수성을 줄이는 등의 효과가 주장되었다. 특히 부분결합을 채용하면 RCE(Reduced Coulomb Energy)망[10] 등과 같이 식별곡면의 구성에 있어서 자유도를 추가할 수 있게 되는데, 이에는 적절한 학습과정이 제시되어야 한다. RCE나 PNN[11] 계열은 본질적으로 망의 외부에서 얻은 지식을 회로화하는 것으로 볼 수 있고, 따라서 단독으로는 학습기능을 가지는 신경망으로 보기 어려운데 반하여 본 연구의 PCR과 PCE망에서는 무작위 결합과 오류역전파 기법을 결합하여 다층퍼셉트론과 같은 유연성을 가진 전형적 신경망을 구성하게 된다.

PCR망을 유한 상태 문법과 궤적발생에 적용한 결과 시계열정보의 처리능력이 매우 뛰어나다는 평가를 얻었으나, 본 연구에서는 인식 대상을 단일 음소로 한정하였으며 상태층이 어떤 궤적, 즉 시이퀀스를 발생하는 것이 아니라, 고정된 결과에 수렴되어야 하므로 이 상태층의 효과가 입력 및 문맥층의 효과를 압도해버릴 가능성이 크다고 판단되어 이 부분을 제외시키고 그림 1의 구조를 택하였다. 비교를 위하여 PCR망에 대한 실험도 병행하였다.

2. 학 습

부분결합 회귀 신경회로망의 학습에는 Rumelhart 등에 의해 개발된 오류역전파(EBP)학습 알고리즘[12]이 이용되었다. EBP는 회망출력과 실제출력간의 오차를 줄이는 방향으로 노드간 연결강도를 수정해 가는 반복학습 과정으로 경사법(gradient descent method)과 유사한 성격을 가지는 것으로 알려져 있다. 패턴 P에 대하여, 학습시점 n일 때 하위층 노드 k와 상위층 노드 q간의 연결강도(weighting factor)를 $W_{pq,k}(n)$이라 하면, 학습에 따른 변화는 다음과 같이 된다.

$$W_{pq,k}(n+1) = W_{pq,k}(n) + \Delta W_{pq,k}(n+1) \quad (1)$$

$$\Delta W_{pq,k}(n+1) = \eta(\delta_{q,k} OUT_{p,j}) + \alpha[\Delta W_{pq,k}(n)] \quad (2)$$

여기서

$OUT_{p,j}$: 패턴 P에 대한 노드 j의 출력값
η : 학습률(learning rate)
α : 관성항(momentum)

이다. 그리고 회로망의 전향연결은 위와 같은 역전파과정에 의해 학습을 받으나 귀환연결과 결정상태층의 자기귀환루프는 역전파 학습을 받지 않는다.

위 식에서 관성항 α는 전번 단계에서 구해진 연결강도의 변화경향을 다음 단계에 반영하는 역할을 하는데 연결가중치 공간에서 오차면(error-surface)의 고주파 변량(high-frequency variations)을 효과적으로 여과한다. 그리고 학습률 η값이 클수록 연결강도의 변화량이 커지게 되므로 학습속도는 증가하지만 수렴하지 못하고 진동할 우려가 있다. 반면에 값이 너무 작을 경우 진동은 없지만 학습속도가 늦어진다.

오류역전파 알고리즘은 비회귀구조를 가진 다층 전향 신경회로망의 학습알고리즘으로 개발되었으나 회귀망에도 이를 또한 적용할 수 있다. 그러나 회귀망에서 역전파 과정이 적절하게 진행되기 위해서는 회귀루프를 가지는 처리요소가 이전 시간스텝에서의 활성을 보존하도록 조치되어야 한다. 본 연구에서 사용된 구조에서는 은닉층과 출력층의 활성상태가 내부상태층과 결정상태층에 각각 보존되고 역전파 과정은 출력층-은닉층-입력층+내부상태층의 순서로 다층퍼셉트론에서와 같이 진행된다.

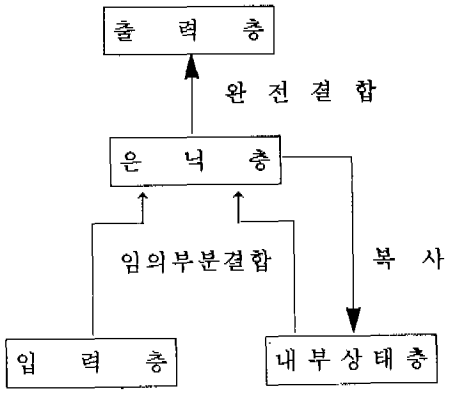


그림 1. 부분결합 Elman망(PCE망)

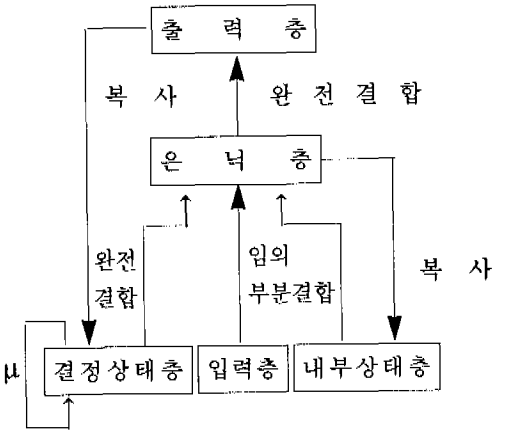


그림 2. 부분결합 회귀망(PCR망)

III. 실 험

1. 음성특징의 추출

관심의 초점은 시간적 변화를 수반하는 음소의 판별에 두었다. 대상으로 ㄱ, ㄷ, ㅂ 3개 자음을 택하였으며 모음들 사이에 나타나는 천이구간적 성격의 표본을 사용하였다. 즉, '사감'에서 '--ㅏ ㄱ ㅏ--', '다도해'에서 '--ㅏ ㄷ ㅗ --', '자본'에서 '--ㅏ ㅂ ㅗ--' 등과 같이 VCV 배열의 자음부분이 인식의 대상이었다.

가. 음성데이터베이스 : 음성데이터는 1

명의 남자화자(speaker)가 150개의 단어를 2회 읽은 것을 바탕으로 하였다. A/D 변환은 DT2801-A를 사용하여 20kHz 샘플링율, 12bit 정밀도로 수행하였고, 이를 8차 저역통과 여파기(차단주파수 4.8kHz)를 통과시켜 10kHz로 샘플링율을 낮추어 기본 데이터베이스를 구성하였다.

나. 시료윈도우 설정 : 앞의 데이터베이스에서 실험에 사용할 VCV 구조의 ㄱ, ㄷ 및 ㅂ부분을 추출하는 것은 수작업으로 하였다. TMS320C30 EVM카드와 'hyper-signal workstation'을 이용하여 파형과 소리를 기초로 160ms정도의 분석구간을 추출하고 역시 수작업에 의하여 자음부의 시작점을 기록하였다. 이 위치는 대단히 불확실하며 여러가지 보조수단이 사용될 수 있으나 본 실험에서는 천이구간적 성격의 부분 전체를 분석의 대상으로 하는 것이므로 모음 이후 에너지가 작아지고 과도적 파형이 들어나는 부분을 대략 선정하는 방식을 택하였다.(그림 3) 이렇게 얻은 각 샘플은 ㄱ-카테고리, ㄷ-카테고리, ㅂ-카테고리의 세 그룹에 대하여 각각 두벌씩 만들어졌으며, 한벌은 학습에, 다른 한 벌은 시험에 사용되었다.

다. 프레임 벡터 : 음성의 특징(feature)으로는 주파수 스펙트럼에 기초한 16차원 벡터를 사용하였다. 한 개의 분석 프레임은 15ms로 하였으며, 먼저 해밍창을 거쳐 1024-점 FFT를 얻고 그 결과를 mel-척도에 기초한 16-채널로 분배하였다. 사용된 필터뱅크 채널의 규격은 표 1과 같으며 주파수 변환공식은 다음을 사용하였다.

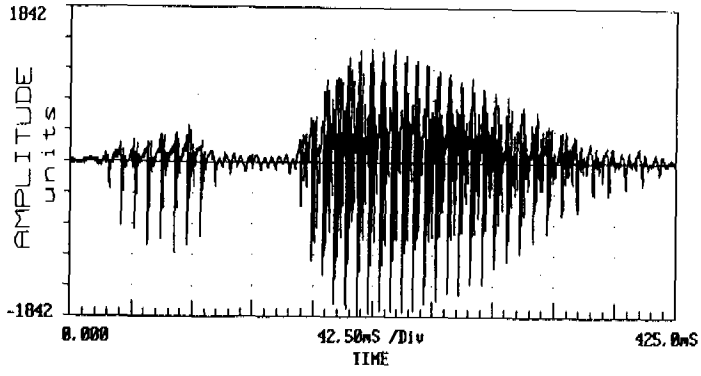
$$z = 26.81 * f/(1960+f)-0.53 \qquad (3)$$

여기서 f는 주파수, z는 Bark로 표현되는 mel-척도 주파수이다.[14] 필터뱅크 출력은 다시 정규화과정을 거쳐서 최종 입력벡터로 하였다. 정규화는 프레임 내의 에너지를 기준으로 하여 프레임 벡터의 제곱을 1이 되도록 하였다.

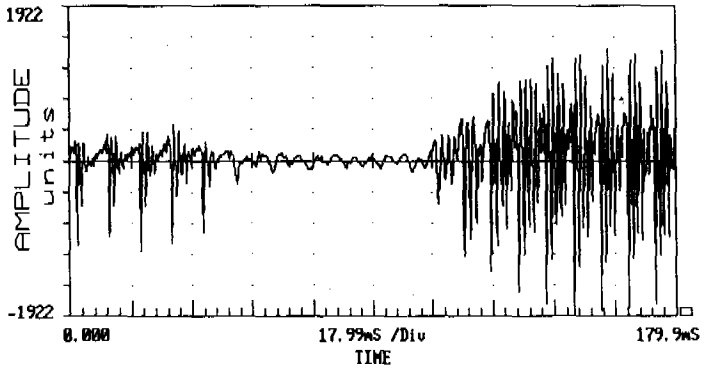
표. 1 필터뱅크의 규격

channel	lower frequency	upper frequency
1	200Hz	297Hz
2	297Hz	404Hz
3	404Hz	521Hz
4	521Hz	650Hz
5	650Hz	794Hz
6	794Hz	954Hz
7	954Hz	1134Hz
8	1134Hz	1339Hz
9	1339Hz	1571Hz
10	1571Hz	1840Hz
11	1840Hz	2152Hz
12	2152Hz	2520Hz
13	2520Hz	2960Hz
14	2960Hz	3496Hz
15	3496Hz	4164Hz
16	4164Hz	5000Hz

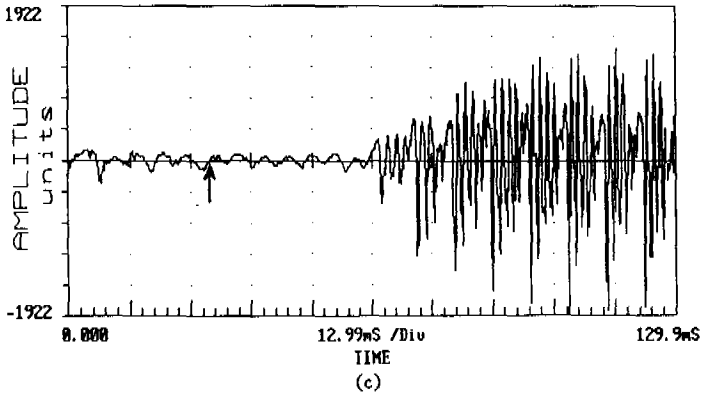
라. 입력시이퀀스 구성 : 학습과 시험에서 사용되는 한 개의 입력 패턴은 앞 항에서 구한 프레임 벡터의 시이퀀스로 구성된다. 프레임간의 간격은 5ms로 하였으며, 학습용으로는 자음부의 시점으로부터 12개 프레임이 사용된다. 즉 70ms 구간이 되어 천이구간과 대략 일치하고 있다. 시험에서는 학습시이퀀스 앞뒤로 30ms 씩을 추가하여 24개 프레임이 사용되었다. 이렇게 하여 시스템 응답의 시간적 변화를 추적해 보고자 하였다. 그림 4는 그림 3.c의 파형에 대한 스펙트럼 및 입력시이퀀스를 보인 것이다. 그림 4.b의 입력시이퀀스에서 가장 큰 면적의 검은 사각형(■)이 최대 스펙트럼값을 나타내고 가장 큰 면적의 흰 사각형(□)이 최소 스펙트럼 값을 나타낸다.



(a)



(b)



(c)

그림. 3 시료윈도우의 선정

(a) 단어 전체의 파형 - '가방'

(b) 해당자음음소로 추출한 파형 - 'ㅏ-ㅂ-ㅏ'

(c) 'ㅂ'을 위한 윈도우내 파형, ↑표는 분석구간의 시작점

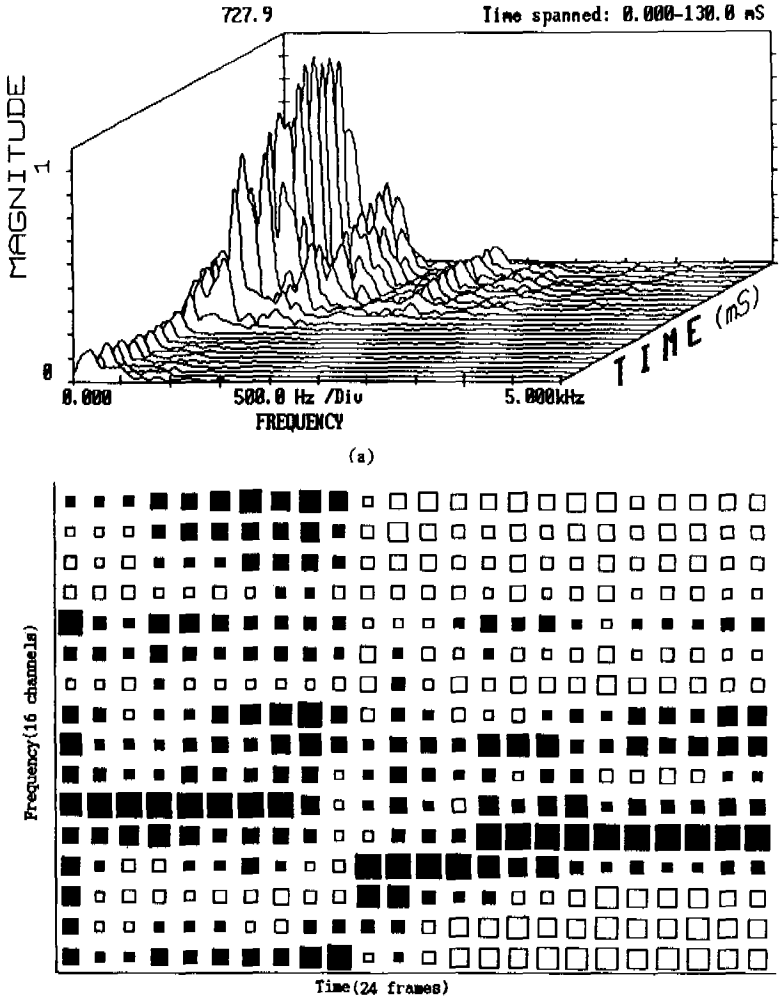


그림 4. a) 그림 3.c의 3차원 스펙트럼
b) a)로부터 발생한 16채널 입력시이퀀스

2. 망 토폴로지

실험에 사용된 망의 구성은 출력층 노드수 3개, 은닉층의 노드수는 80개로 하였으며, 내부상태층 역시 80개 노드로 구성하였고 입력층의 요소수는 16개로 하였다. 또한 입력층 및 내부상태층과 은닉층 사이의 부분결합 밀도(Tuple수)는 각각 16개, 10개로 구성하였으며, 학습에 사용된 학습률(learning rate)과 모멘텀(momentum)은 실험을 통하여 각각 $\eta=0.05$, $\alpha=0.9$로 하였다. 병행하여 시험된 그림 2의 PCR망의 구성에서는 출력층과 같은 상태층이 추가되었다. 그림 5는 수렴과정을 보이고 있다. 8,000번 정도의 반복으로 충분히 수렴됨을 알 수 있었으며 SUN-SPARC 웍스테이션으로 각각 15개의 샘플로 구성된 세 개의 카테고리를 학습하는 데 약 20시간이 걸렸다.

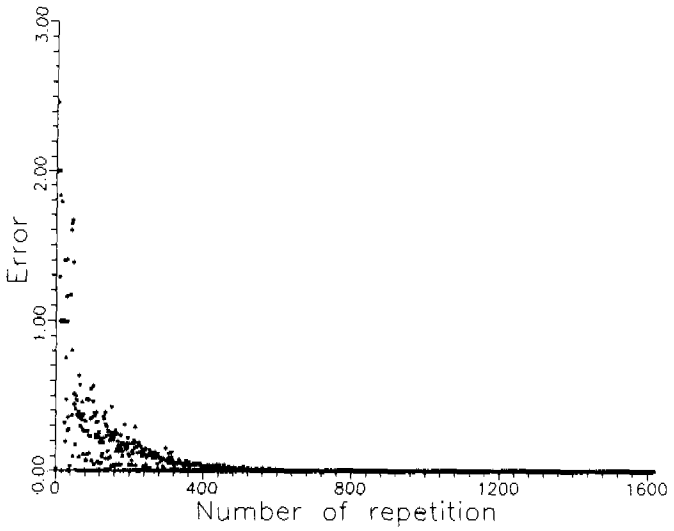


그림 5. 학습 횟수에 따른 수렴과정

IV. 토 론

1. 인식률

표 2는 인식 시험결과를 비교한 것이다. 여기서 인식의 판정은 입력시이퀀스의 전기간동안 출력층 활성도를 누적하여 최대가 되는 노드를 선택한 것이다. PCE와 PCR망 모두 시험되었는데 학습에 사용한 샘플에 대해서는 두 망 모두 100% 인식을 나타내고 있다. 학습에서 제외되었던 시험패턴에 대해서는 PCR망은 평균 62.2%, PCE망은 73.3%의 인식을 보이고 있다. PCR망의 출력층 활성도를 추적해 본 결과 거의 처음에 나타난 경향이 끝까지 계속되는 형태를 보였다. 이것은 상태층의 효과가 너무 강하게 작용하도록 교육되는 것을 의미하며 초기 입력 프레임이 최종결과를 좌우하는 결점을 가진다. 상태층을 제거한 PCE망은 이런 결점을 완화하는 것을 나타내었으며 인식률도 개선되었다.

본 연구의 결과는 TDNN의 인식률과 비교하면 상당히 낮음을 볼 수 있다. TDNN이 가지는 본질적인 우수성 외에도 다음과 같은 추가적 요소가 작용하고 있다고 생각된다. 1) TDNN의 규모는 본 연구의 것과 비교할 때 처리요소, 즉 연결의 수만으로 비교해도 대략 3배의 것이며 여기에 대량의 학습데이터를 수용하면 일반화의 여유가 상당히 클것으로 예상된다. 2) TDNN에 적용한 데이타는 기본적으로 잘 예비된 초성의 자음을 활용하고 있어 특징이 비교적 두드러진 시료라고 할 수 있다. 이에 비하여 본 연구에서는 모음 사이에 찰라적으로 지나가는 과도현상과 같은 표본들만 대상으로 하였으므로 포착이 매우 어렵게 되어 있다. 3) TDNN에서 사용되고 있는 입력특징벡터는 대단히 주의깊게 준비되어 있다. 16채널 필터뱅크로 구성된 15개의 연속적 패턴을 활용한다는 점은 유사하다. 그 다음 전체 에너지를 기준으로 ±1의 범위로 재구성하고 있는데 이것은 인식의 효과를 올리기 위한 방편으로 이해되나 논리적 당위성에는 다소 문제점이 있다. 첫째로 에너지의 정규화는 프레임 단위가 아닌 블럭단위로 하고 있다. 이것은 시계열 정보처리에 있어 매우 바람직한 방법이지만 구간을 설정하기 위하여 미리 적절한 분절(segmentation)이 이

표 2. 두가지 망에 대한 인식율 비교

망 종 류	PCE망				PCR망			
음 소	ㄱ	ㄷ	ㅂ	종합	ㄱ	ㄷ	ㅂ	종합
학 습 패 턴	15/15	15/15	15/15	45/45 (100%)	15/15	15/15	15/15	45/45 (100%)
시 험 패 턴	10/15	13/15	10/15	33/45 (73.3%)	6/15	8/15	14/15	28/45 (62.2%)
오 류 내 용	ㄷ:1	ㄱ:1	ㄱ:5	12/45	ㄷ:4	ㄱ:4	ㄱ:1	17/45
	ㅂ:4	ㅂ:1	ㄷ:0	(26.7%)	ㅂ:5	ㅂ:3	ㄷ:0	(37.8%)

루어져야 한다. 둘째로 스펙트럼의 낮은 에너지 부분은 논리적으로 효과가 작아야 하는데 여기서는 -1이 되므로 오히려 매우 커지는 모순이 있다. 본 연구에서도 이와 같은 주변요소의 효과를 고려해서 입력을 조율하면 인식율은 어느정도 제고될 것으로 예상되지만 관심의 초점이 시계열에 관한 속성의 평가에 있었으므로 현 상태에서 분석하였으며 인식 시스템의 개발단계에 따라 추가적 신호처리 과정이 포함될 수 있을 것이다.

2. 출력노드 활성의 시간적 변화

그림 6에는 출력노드의 활성패턴을 몇 개 보였다. 이 패턴은 24프레임의 입력을 차례로 공급하고 출력의 변화를 관찰한 것이며, 이에 대한 고찰을 정리하면 다음과 같다. 1) 앞 절에서 지적한 바와 같이 PCR망의 경우에는 초기부터 강력한 결론에 수렴하는 경향이 있었는데 PCE망에서도 이런 경향이 눈에 띈다. 그림 6.a의 경우, 24프레임 중 앞뒤 6개씩 12프레임은 자음부가 아닌 전후 모음부 임에도 불구하고 처음부터 결론이 나타난다. 시이퀀스의 파악능력에 의한다기보다 원래 학습패턴에 포함된 유사 프레임 패턴에 의한 판정의 경향을 보인다고 할 수 있으며 적절한 수준으로 통제될 필요가 있다. 2) 그림에서 보는 바와 같이 3개의 노드중 1개가 활성화되면 다른 노드가 억제되는 인상을 받는다. 즉 'winner takes all' 의경쟁학습의 경향을 나타낸다. 그러나 망구조를 검토해 볼 때 실제로는 강력한 분류기능의 결과가 그와 같은 착각을 일으키는 것으로 판단된다. 입력층은 다수의 은닉층에 부분결합으로 연결되어 있다. 따라서 식별곡면은 연속적이며 볼록형(convex)구간일 필요가 없이 부분적인 집합으로 자유롭게 구성이 가능하다. 완전결합의 경우 입력 카테고리간 함수적 독립성을 유지할 수 없을 때에는 상호 영향을 줄 수 있는데, 이에 비하여 뚜렷한 강점을 드러내고 있다. 3) 0 또는 1에 가까운 값 외에 중간적 값이 잘 드러나지 않고 있다. 이것은 2)항의 강력한 분류기능과 1)항의 프레임별(시이퀀스가 아닌) 학습들과 결합된 것으로 보이며, 보기에는 좋은 특성 같지만 시이퀀스 학습에 취약점이 있음을 암시한다.

V. 결 론

본 연구에서는 부분결합 및 회귀형의 동적신경망 모형을 음소 인식에 적용하고자

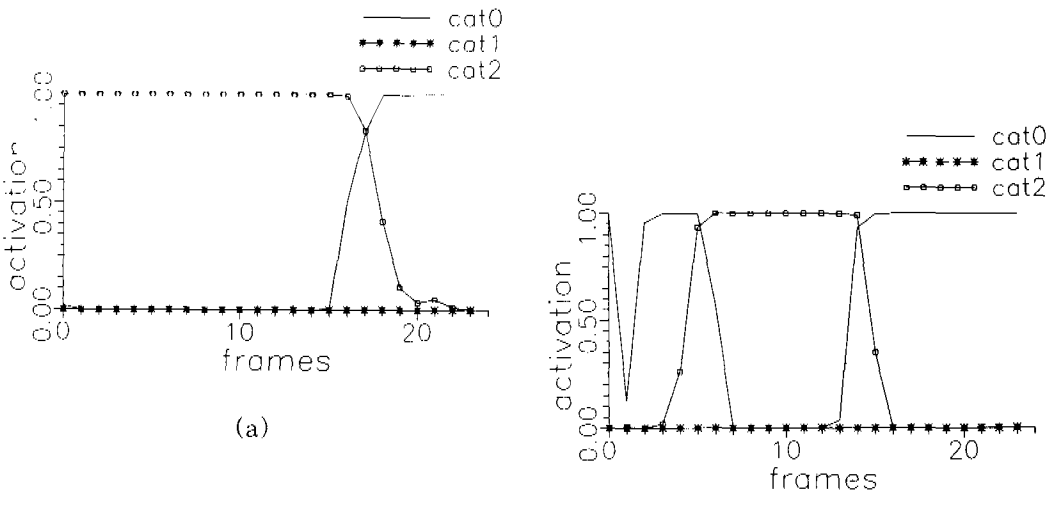


그림 6. 출력노드 활성의 시간적 변화
(a) 13번째 샘플의 활성결과
(b) 첫번째 샘플의 활성결과
(a)(b) 모두 ㅂ-카테고리로 인식

시도하였다. 인식률은 73.3%로서 기존의 시간지연망에 비교하여 많이 뒤지는 것이었으나 시간지연망의 엄청난 규모와 입력형태의 조절등 인식률 향상을 위한 각종 주변작업을 고려하면 그 차이는 중요한 것이 아니며 오히려 빠른 수렴, 강력한 판정 등 고무적인 경향을 인식할 수 있었다. 그러나 앞의 토론에서 드러난 바에 의하면 실험된 망구조는 본질적 취약점이 있음이 지적되며 이를 근거로 새로운 방향을 제시하고자 한다. 시간지연망에 대한 본 연구자의 비판은 비록 인식능력은 양호하지만 시간지연망은 본질적으로 시간적 변화특성을 인지한다기보다는 시간-주파수 2차원 특징을 시간축상에서 수없이 반복 변위하여 교육함으로서 어느 위치에서도 감지할 수 있도록 한다는 개념이 강하다는 것이다. 한편, 본 연구의 회귀망은 논리적으로는 시간적 변화를 수용하도록 되어있으나, 토론에서 밝혀진 바와 같이, 인식률 면에서 다소 우월한 것으로 나타난 PCE망은 시이퀀스보다는 단위 프레임에 영향을 더욱 받는 것으로 보인다. 따라서 HMM과 같은 강력한 동적모형을 구현하기 위해서는 시이퀀스 처리 기능이 더 우수한 PCR망을 사용하여 상태천이 기능을 활용할 필요가 있다. 여기서 문제가 되는 것은 현재의 학습방식이 출력층의 단일노드, 즉 ㄱ, ㄷ, ㅂ 중 어느것이냐를 판정하는 것으로 되어 있어서 통상의 오류역전파 계통에서와 마찬가지로 출력층의 상태를 고정하여 교육하는 형태로 되며 천이패턴을 발생하거나 추적할 필요가 없게 되어있다. 이러한 상황에서는 PCR 망은 앞에서 언급한 바와 같이 상태층의 영향력이 너무 커지는 결과를 초래하여 최초의 프레임의 영향이 지배적이 될 가능성이 높다. 이를 완화하는 목적으로 회망출력의 값을 0으로부터 1까지 점진적으로 증가시켜가며 교육함으로써 입력 시이퀀스의 초기에 상태층이 포화되는 것을 방지하는 방법을 시도하여 보았

다. 이때에는 망이 수렴에 이르지 못하는 것을 발견하였는데 그 원인은 입력샘플의 순차구조를 처리할 수 없는 상황에서 같은 입력패턴에 대하여 시간에 따라 다른 출력값으로(0.1, 0.2 등으로)학습시키기 때문에 생기는 혼란이라고 생각된다. 본 연구의 결과를 종합하여 보면, 최종판정출력과 은닉층 사이에 상태층을 삽입하는 새로운 PCR 망을 구성하고, 이에 적절한 학습과정을 정의함으로써, 상태표현 및 그 변화를 허용할 수 있는 기능을 부여할 수 있을 것으로 생각하며, 이러한 접근은 보다 논리적 당위성을 포함시킬 수 있을 것으로 기대한다.

참고문헌

1. A. Waibel and K-F. Lee, ed., Readings in Speech Recognition, pp.2, San Mateo CA : Morgan Kaufmann Publishers, 1990.
2. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-26, pp.43-49, Feb. 1978.
3. L.R. Rabiner and B.H. Juang, "An Introduction to Hidden markov Models," IEEE ASSP Magazine pp. 4-16, Jan. 1986.
4. 이태호, "음성인식을 위한 인공 신경망", 컴퓨터기술, 제6권 제1호, pp.45-49, 1989.6.
5. R.P. Lipmann, "An Introduction to Computing with Neural Nets," IEEE ASSP Magazin, pp.4-22, Apr. 1987.
6. J.L. Elman and D. Zipser, "Learning the Hidden Structure of Speech," Tech. Rep. ICS-8701, Univ. Calif., San Diego, Feb. 1987.
7. T. Kohonen, "The Neural Phonetic Typewriter," IEEE Computer, pp. 11-22, March 1988.
8. S. S. Kim, et al, "Learning Sequential Structures in Partially Connecterd Recurrent Neural Networks," KITE Journal of Electronics Engineering, vol.2, no. 1, pp.109-115, June 1991.
9. F. Tsung and G. W. Cottrell, "A Sequential Adder Using Recurrent Networks," Proc. IJCNN'89, IEEE/INNS, vol.2, pp.133-139, June 1989.
10. B. Widrow, study director, DARPA Neural Network Study, pp.90 AFCEA International Press, 1987.
11. D. F. Specht, "Probabilistic Neural Networks," Neural Networks, vol. 2, pp.109-118, 1990.
12. D.E. Rumelhart and J.L. McClelland, Parallel Distributed Processing: Experiments in the Microstructure if Cognition, vol. I, pp.318-362, Cambridge, MA: M.I.T. Press, 1986.