

복합신경망을 이용한 단음인식에 대한 연구

A Study on the Single Word Recognition using Composite Neural Network

김 석 동*, 이 행 세**
(Seok Dong Kim, Haeng Se Lee)

요 약

본 논문은 신경망을 이용한 음성인식에 관한 연구이다. 우선 음성을 음소로 분리하였다. 그런다음 각각의 음소를 인식하였다. 음소인식을 위해 제어망과 몇개의 소규모 망으로 구성된 복합신경망을 이용하였다. 제어망은 음소가 어느집단에 속하는가를 결정하고, 소규모망에서는 각 집단에 속하는 음소를 인식한다. 마지막으로 프레임별 인식 결과를 누적하여 음소를 판별하였다.

ABSTRACT

This paper is a study on the speech recognition using neural network. First, the phoneme is separated from speech by neural network. Then each phoneme is recognized by another neural network in frame unit. For the purpose of phoneme recognition, the composite neural network which consists of a control network and several sub-networks is used. The control network identifies the group to which the phoneme belongs and the sub-network recognizes the phoneme in each group. Finally, phoneme is discriminated by intergating the result of recognition for each frame.

I. 서 론

음성은 사람들 사이의 대화에서 가장 자연스러운 정보전달수단이다. 만일 기계가 사람의 음성을 정확하고 신속하게 인식할 수 있다면 여러분야에서 유용하게 이용될 수 있을 것이다. 음성을 자동적으로 표기할 수 있고, 사람과 기계사이의 간단한 대화가 가능하며, 청각기능을 상실한 사람들에게 도움을 줄수

가 있을 것이다. 그러나 아직까지 오늘날의 음성인식 기로는 다양한 사람의 연속적인 음성을 인식하기에는 해결해야할 많은 문제점을 갖고있다. 사람에 따라 음성은 특유의 성질을 갖고 있으며, 시간에 따라 수시로 변화하는 것이 대표적인 문제로 볼 수 있다^[1].^[2] 어떤 사람은 보통 보다 빠르고/늦게 발음하거나, 크고/작게 말을 하며, 심자어는 같은 사람이라 해도 똑같은 단어를 동일한 형태로 발음하지 않는다^{[1][3]}. 더구나 음성인식에서 올바르게 인식이 되었는지 결정하기도 쉽지가 않다. 일반적으로 사람의 경우 듣는 순간 5만개의 단어와 순간적으로 비교해가면서 음성

*호서대학교 전자계산학과 조교수

**아주대학교 전자공학과 교수

접수일자: 1992년 8월 15일

학적이나 구분적이나 의미적인 다양한 분석을 행하면서 인식하고 있으나⁴⁾ 아직 음성인식기는 단순한 형태의 분석에 의해 인식되고 있는 실정이다.

지금까지 가장 일반적인 인식알고리즘으로서는 DTW(Dynamic Time Wrapping)이나 HMM(Hidden Markov Model)이 있다. DTW는 저장되어 있는 표준패턴과 미지의 음성을 동적인 프로그램 기법을 이용하여 정합시키는 방법으로 시간축상에서 표준패턴의 각부분과 정합시켜 가장 비슷한 것을 찾는 방법이다⁵⁾. 현재까지 가장 널리 사용되는 음성 인식으로 HMM을 들 수 있다. 음성을 인식하는데 통계적으로 접근하는 방법인 HMM은 음성데이터를 관찰열로 사용하여 HMM에서 만들어진 여러형태의 관찰열 중에서 가장 유사한 것을 선택하는 알고리즘이다⁶⁾. 근래에 와서 인간의 두뇌를 모델링한 신경망에 대한 연구가 활발해지면서 음소와 음절을 기본으로 하는 인식을 신경망에 의해 행하여 지고 있다. 음소인식에 대한 가장 성공적인 신경망은 시간지연을 이용한 신경망(TDNN)이다⁷⁾. TDNN은 여러층을 갖는 오류 역전파(Back Propagation) 모델을 이용하여 각각의 연속적인 층으로 오랜시간에 대하여 음성을 적용하는 방법으로 가장 위층에서 시간에 불변하는 특성을 가진 음소를 인식하는 방법이다⁸⁾.

본 논문에서는 우선 인식하려는 음성을 음소로 나누고, 각각의 음소를 인식하기 위하여 여러개의 신경망을 복합적으로 사용하였다. 우선 단음을 음소로 분리하기 위하여 음소분할 신경망을 적용하고, 분리된 음소를 다시 소규모로 구성된 여러개의 신경망에 의해 인식하는 방법을 사용하였다. 신경망에 사용한 알고리즘으로는 오류 역전파(BP)알고리즘을 이용하였다⁹⁾. 음성인식분야에서 적은수의 어휘를 인식하는데 BP망은 높은 인식률을 얻을 수 있다고 발표되어 있다¹⁰⁾. 그러나 어휘가 많은 음성을 일반적인 BP망으로 적용하면 훈련시간이 망의 크기에 비례하여 매우 커지게 된다는 단점이 있다. 더구나 커다라 망내에서는 해(solution)공간내의 총체적인 최소점보다는 부분적인 최소점을 발견할 확률이 높아 최적의 해를 구하기가 매우 어렵다. 또한 새로운 어휘를 추가할때 전체적인 망을 재훈련시켜야 하는 반복적인 문제가 발생한다. 그러므로 일반적인 BP망으로는 많은 어휘에 대해서는 실효성이 없게 된다. 이러한 문제점

을 보완하기 위해서 본 논문에서는 전체 인식대상을 몇개의 그룹으로 나누어 훈련시키는 방법을 사용하였다¹¹⁾. 부분망은 그 그룹에 속하는 데이터만을 훈련시키므로서 훈련시간을 감소시킬 수 있으며 새로운 어휘에 대해서도 손쉽게 확장할 수 있는 장점을 가진다.

음성인식을 위한 특징 파라미터 추출 방법으로는 시간영역 분석방법¹²⁾을 이용하였고, 자음을 훈련시키기 위하여 자음 부분만을 추출하기 위한 자·모 분리는 자기상관과 에너지넷 영교차율을 이용하였다.

II. 학습 데이터 선정

음성을 신경망에 의하여 인식하기 위해서는 정확한 학습 데이터가 필요하다. 본 논문에서의 인식대상은 자음과 모음이 결합되어있는 음성이므로 학습데이터인 모음은 단모음중 /아/, /어/, /오/, /우/, /으/, /이/, /에/를 사용하였다. 단모음중 /에/와 /애/는 정확한 발음을 얻기가 어려워 자주 사용되는 /에/를 택하였다. 자음은 독립적으로 발음할 수 없으므로 모음과 같이 발음한 음성에서 추출하였다. 자음부분 추출방법은 크게 두부분으로 구성된다. 첫번째로는 단구간 자기상관(short-time autocorrelation)을 이용하여 잡음성이 강한 자음(무성자음)이 포함된 음성과 모음성이 강한 자음(유성자음)이 포함된 음성을 구분하고, 두번째로는 무성자음부분은 단구간 영교차율을 이용하고 유성자음부분은 단구간 에너지를 이용하여 자음부분과 모음부분을 분리하였다. 본 논문에서 사용한 유성자음과 무성자음의 분류에 대한 내용을 간략히 정리하면 다음과 같다. 음성신호를 $x(m)$ 이라 할때 단구간 자기상관 함수 $\phi(k)$ 는

$$\phi(k) = \sum_{m=0}^{N-k} x(m)W(1-m)x(m+k)W(1-m-k)$$

가 된다. 여기서 1은 프레임을 나타내고 k는 지연시간이고 N은 프레임의 길이이며 W는 창(Window)함수이다. 본 논문에서 사용한 창함수는 해밍창으로

$$W(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (N-1)), & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$

이다. 1번째 프레임에서의 $\phi(0)$ 를 제외한 최대 자기 상관값을 r_1 이라 한다.

$$r_1 = \operatorname{argmax}_{k=1 \dots N-1} \{\phi(k)\}$$

이다. 여기서 $\operatorname{argmax}\{\}$ 는 괄호내의 값중 최대가 되는 것을 선택하는 함수이다. 일반적으로 유성음과 무성음에 따라 r_1 값이 다르다. 일반적으로 에너지와 영교차율은 작고 r_1 이 클때는 유성자음이고, 에너지와 r_1 은 작고 영교차율이 클때는 무성음임을 알 수 있다 [11].

본 논문에서 사용한 자·모음의 결정방법은 일종의 패턴인식방법으로 자음과 모음사이의 음성학적 특징을 이용하였다. 첫째로 무성자음과 모음의 분리방법은 다음과 같다. 파열음이나 마찰음과 같은 무성자음은 잡음과 비슷하여 주파수가 높으면서 에너지가 낮다. 이를 이용하여 본 논문에서는 자·모 분리를 우선 고주파 성분의 백색 잡음이나 팝 노이즈와 같은 저주파 성분을 제거하기 위해 음성신호를 band-pass여파기를 통과시키고나서, 자음에서 모음으로 변화되는 천이구간에서의 작은 저주파 성분을 이용하여 자·모음의 분리를 하였다. 둘째로 유성자음과 모음의 분리

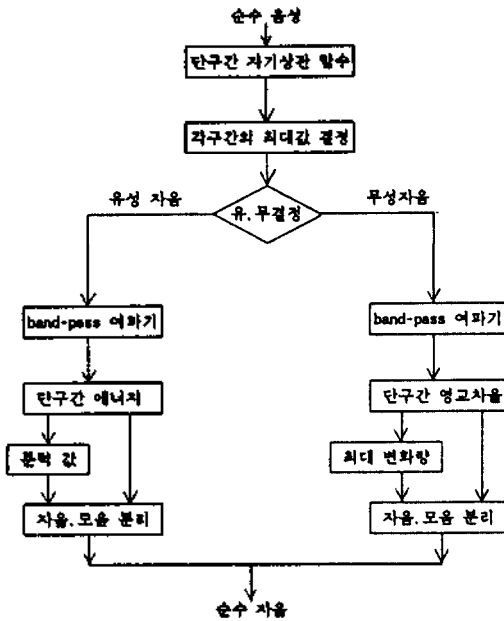


그림 1. 자음 부분 추출 방법

방법은 다음과 같다. 비음과 같은 유성자음은 모음과 비슷한 특징을 가지며 에너지와 주파수가 약간 작다. 음성의 전구간에 대하여 단구간 에너지를 구하면 모음부분에서 최대값을 갖는 프레임이 나타난다. 최대값이 나타나는 프레임에서 시작하여 앞 프레임으로 이동하면서 문턱값(Threshold value)보다 작은 값을 갖는 프레임이 자음부분임을 알 수 있었다. 그림 1에서 자음 부분 추출 방법에 대한 전반적인 알고리즘을 나타내었다.

이와 같은 방법에 의해 추출된 자음은 다시 유사한 특성을 갖는 음성들끼리 나누었다. 즉 /ㄱ, ㄷ, ㅂ/, /ㄴ, ㄹ, ㄴ, /, /ㅅ, 스, ㅈ/, /ㅋ, ㅌ, ㅍ, /, /ㅎ/으로 5그룹으로 나누고 모음까지 총 6개의 그룹으로 분리하여 학습 데이터를 선정하였다.

III. 인식 알고리즘

본 논문에서 제안하는 음성인식은 두가지 종류의

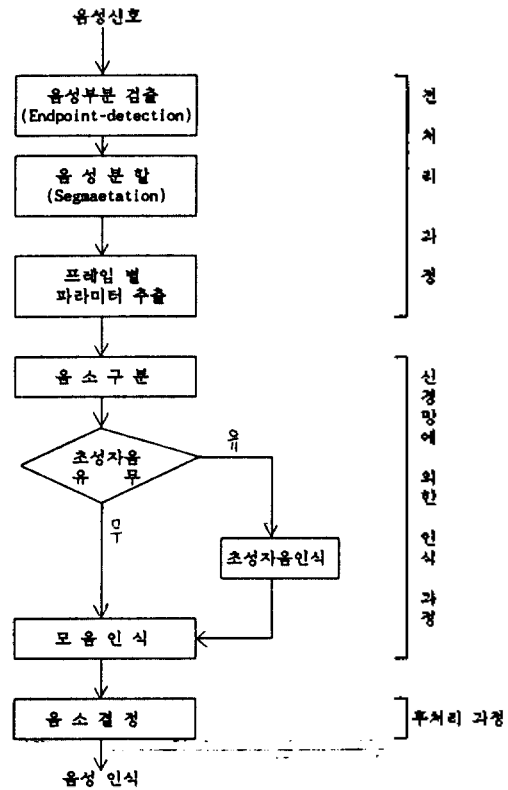


그림 2. 음성인식에 대한 전체 개략도

신경망에 의해 수행된다. 하나는 음성에서 음소를 분류하는 음소구분 신경망과 다른 하나는 분류된 음소를 인식하는 음소 인식 신경망이다. 음소 인식 신경망은 규모가 작은 보통의 신경망이 여러개로 구성하였다. 신경망에 의하여 프레임별로 인식된 결과는 음소결정과정을 거쳐 최종적으로 음소를 결정한다. 그림 2에 인식에 사용한 전체 개략도를 나타내었다.

3-1 음소 구분

본 논문에서의 인식 대상음성은 1)모음음성과 2)초성자음과 모음이 결합된 음성인 두가지 종류이다. 음소구분 신경망에 의해 음성내에 초성자음이 있을 때 자음성분과 모음성분을 분류한다. 그림 3에 자음과 모음의 위치를 결정하는 알고리즘을 나타내었다. 음소 구분 신경망은 음성을 프레임 별로 자음과 모음으로 구분하는 신경망으로 두개의 출력 노드를 갖는다. 학습데이터는 2장에서 얻은 모든 음성을 2개의 그룹으로 나누어 훈련시켰다. 인식결과는 프레임 별로 나타나므로 그림 3b)와 같이 오인식 나타날 수가 있으며 특히 자음에서 모음으로 천이되는 부분에서는 불안정한 상태로 인식이 되어 음소구분이 용이하지 않으므로 비교적 안정한 상태를 만들기위해

Filtering을 하였다. 즉 인식 결과를 한 프레임씩 이동하면서 5프레임중 3개 프레임 이상이 같은 결과가 나타날때를 인정하였다. 그림 3c)에 안정한 형태의 인식결과를 보였다. 자음은 모음보다 짧게 발음이 되므로 전체 프레임 수 (fend)의 1/2 이하의 곳에서 자음과 모음의 구분이 이루어 진다고 볼 수 있으므로 앞 프레임(정확하게 3번째 프레임)부터 fend/2 까지를 조사를 하였다. 현재의 프레임 보다 앞의 2개 프레임과 뒤의 2개 프레임이 각각 같은 값을 가지면서 서로 다른 값을 가질때를 자음과 모음의 구분점 (fcv)으로 결정하였다.

3-2 프레임별 음성인식

음소로 분리된 프레임들을 다시 음성으로 인식하기 위하여 음소 인식 신경망을 사용하였다. 자음과 모음의 구분이 없는 경우(fcv=0)에는 순수 모음음성이므로 모음인식 신경망에 의해 음성을 인식할 수 있고 0 < fcv ≤ (fend/2)이면 모음 앞에 자음이 있는 경우이므로 자음 인식 신경망과 모음인식 신경망에 의해 인식된다. 모음 인식 신경망의 출력은 7개의 모음을 인식하므로 7개의 노드로 구성하였고, 자음은 총 13개를 인식해야 하므로 자음인식 신경망의 규모를 줄이기 위하여 여러개가 결합된 소규모의 신망을 사용하였다.

신경망에 의해 패턴을 인식할 때의 커다란 문제점 중의 하나는 훈련시간을 가급적 적게 하면서 인식률을 높이는 것이다. 훈련시간을 줄이는 방법으로는 여러가지가 있다. 총체적인 오차가 제일 적은 곳 즉 최적해를 찾아가는 수렴 속도를 증가시키는 방법으로 모멘텀을 사용하거나, 훈련할 데이터를 선별하나, 인식할 패턴에 가장 잘맞는 신경망 모델을 선택하는 방법등이 있을 수 있다. 본 논문에서는 훈련시간을 결정하는 훈련횟수를 줄이기 위해 우선 비슷한 특성을 갖는 데이터를 묶어서 여러개의 부분집합으로 나누고, 각 부분집합에 속한 데이터만을 대상을 훈련하는 방법을 사용한다. 그림 4에 자음인식에 사용한 신경망의 구조를 나타내었다. 가장 우측에 있는 신경망(그림에서 B)은 음성이 어느 그룹에 속하는 지를 결정하는데 사용하고, 나머지 신경망(그림에서 A1-A4)은 각기 그룹에 속하는 자음을 인식하는데 사용한다. 각각의 부분 신경망은 각각 개별적으로 학

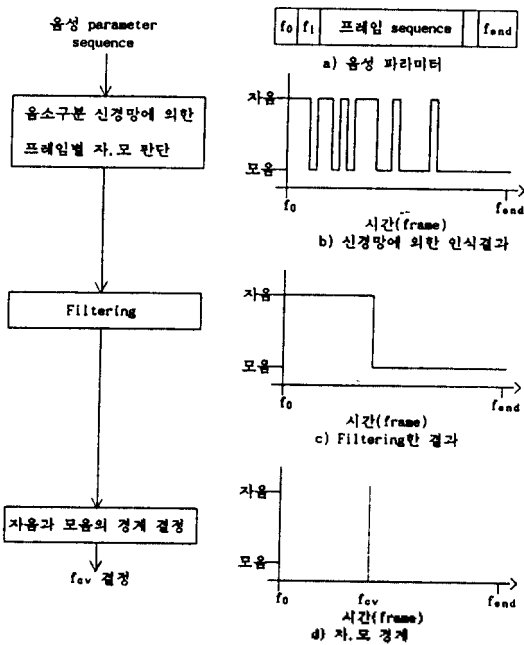


그림 3 음소 구분 알고리즘

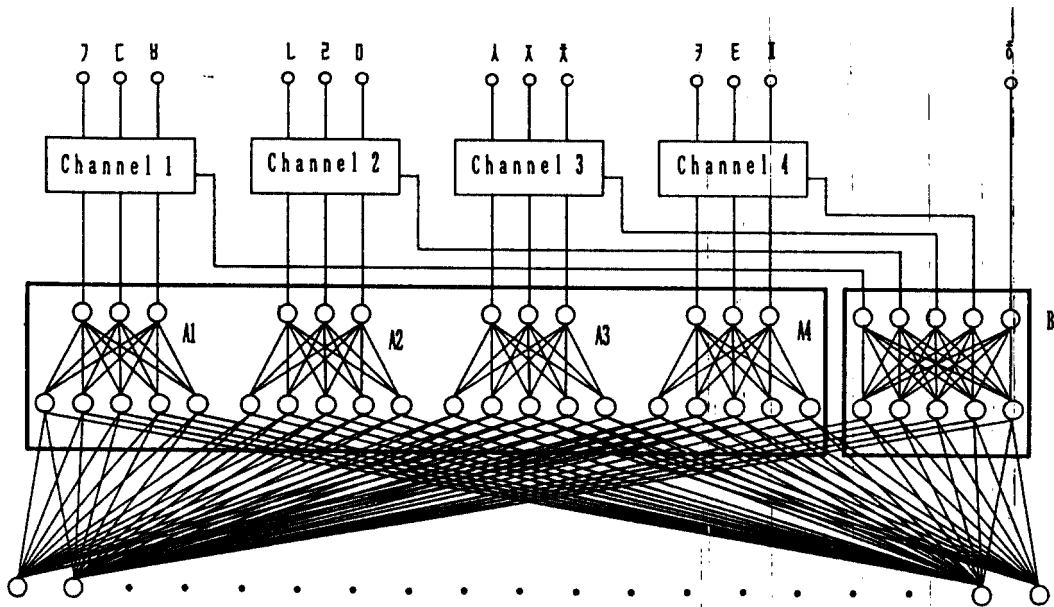


그림 4. 음소인식 과정

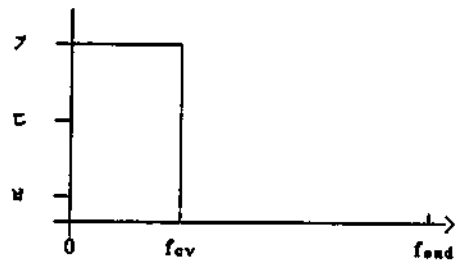
습한다. 그러므로 각부분망의 학습시간은 짧아진다. 최종인식결과는 그룹을 결정하는 신경망의 출력과 그룹별 자음을 인식하는 신경망의 출력에 의해 나타난다. 대상 자음은 5개 그룹으로 나누었으나 그룹에 속하는 자음이 하나밖에 없는 /ㅎ/은 그룹을 결정하는 신경망의 출력이 바로 /ㅎ/의 인식을 의미한다.

3-3 음소 결정

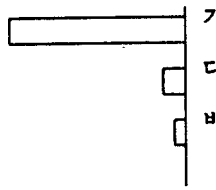
신경망에 의해 프레임 별로 인식된 최종적으로 음성으로 인식하기 위해서 음소의 조합으로 나타내야 한다. 음소 구분 신경망과 자음인식 및 모음인식 신경망등 3종류의 신경망을 거치면서 음소를 결정하는 과정을 종합하면 1)음소 구분 신경망의 결과에 의해 자음과 모음의 위치 (f_{cv})가 결정되며 2)초성자음에 대한 프레임별 인식 신경망의 결과에 의하여 음성의 시작점 부터 f_{cv} 까지 인식결과를 누적하여 최대값을 갖는 음소를 결정하고 3)모음에 대한 프레임별 인식 신경망의 결과에 의하여 f_{cv} 에서 부터 음성의 마지막 (f_{end})까지 인식결과를 누적하여 모음을 결정하게 된다. 그림 5에 음성 /가/에 대한 음소 인식과정을 나

타내었다.

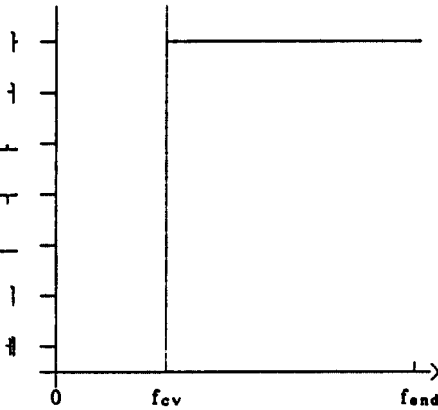
음성 /가/를 우선 자음 인식 신경망에 의해 인식을 하면 그림 5a)와 같이 나타난다. 여기서 자음 인식은 처음 부터 f_{cv} 까지가 유효하므로 그림 5b)에서 보는것 같이 f_{cv} 까지 인식결과를 누적하면 /ㄱ/은 큰 값을 가지고 나머지 /ㄷ/과 /ㅂ/은 작은 값을 알 수가 있다. 이 음성을 다시 모음인식 신경망에 의해 그림 5c)와 같은 결과를 얻게되며 이번에는 f_{cv} 에서 마지막 프레임인 f_{end} 까지 누적하여 그림 5d)와 같은 결과를 얻게 된다. 그림 5c)와 5d)에 의해서 /가/의 음성을 인식할 수 있다.



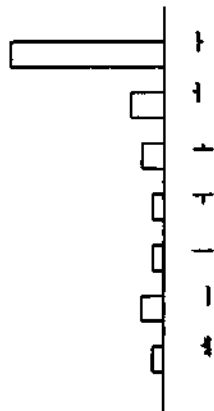
a)음성 /가/에 대한 음소 인식과정을 나타내었다.



b) 자음 인식 누적결과



c) 음성 /가/에 대한 모음인식결과



d) 모음인식 누적 결과

그림 5. 음소인식 과정

IV. 실험 및 결과

음성 데이터 수집은 DAT(digital audio tape recorder)로 방음실에서 다이내믹 마이크를 통해서 약 1.6sec 동안 발음한 음성을 자기 테이프에 저장 하였다. 학습에 사용한 음성은 모음으로는 단모음인 /ㅏ/, /ㅑ/, /ㅓ/, /ㅕ/, /ㅗ/, /ㅛ/, /ㅜ/, /ㅠ/, /ㅡ/, /ㅣ/ 등 7개음성이고, 자음으로는 /ㄱ/에서 부터 /ㅎ/까지

/ㅇ/을 제외한 13개의 자음소에 모음 /ㅏ/와 /ㅑ/ /의 결합 형태인 총 26개의 음성 데이터를 사용하였다.

인식에 사용한 음성은 학습에 참여하지 않은 사람의 음성으로 /ㅇ/을 포함한 자음소 14개와 단모음 7개가 결합한 총 14×7=98개의 음성을 대상으로 하였다. 자기 테이프 재생기가 증폭기를 거쳐 나온 음성을 대상으로 하였다. 자기 테이프 재생기와 증폭기를 거쳐 나온 음성을 Data Acquisition Board인 Data Translation 2801 Board로 음성 크기는 최대치가 ±10V 이내, 샘플링 주파수 10kHz, 양자화 준위 12bit로 아날로그 디지털 변환(ADC, analog digital conversion) 하여 저장 하였다. 신경 회로망의 입력 노드 수는 15차의 PARCOR 계수를 사용하였다. 프레임의 길이는 10ms로 하였고, 자음 구간의 분석 구간을 늘리기 위해 6ms를 중첩 시켰다. 중간 노드의 수는 여러번의 반복적인 실험에 의해 신경망의 종류에 따라 각각 따로 정하였고, 출력 노드의 수도 마찬가지로 결정하였다. 학습 횟수는 시간을 제한하기 위해 2,000회 이내에서 각 음성에 대해 실제의 출력값과 원하는 출력값의 차 즉 허용 오차가 0.3이내에 들도록 하였다. 실험에서 사용한 음성데이터 내용과 특징과라미터를 표 1에 나타내었다.

표 1. 음성 데이터 채집과 특징 파라미터

대상 자음	/ㄱ/, /ㅋ/, /ㆁ/, /ㄴ/, /ㄷ/, /ㄹ/, /ㅁ/, /ㅂ/, /ㅅ/, /ㅇ/, /ㅈ/, /ㅊ/, /ㅌ/, /ㅍ/, /ㅊ/, /ㅎ/
대상 모음	/ㅏ/, /ㅑ/, /ㅓ/, /ㅕ/, /ㅗ/, /ㅛ/, /ㅜ/, /ㅠ/, /ㅡ/, /ㅣ/
음성 채집	성인 남자의 총 98개 음성
샘플링 특성	샘플링 주파수: 10kHz 양자화 비트수: 12bit
분석 구간	프레임 길이: 10ms, 프레임 중첩: 6ms 해밍창 사용
특징 파라미터	부분 자기 상관 계수(15차)

표2는 자음을 인식하기 위해 사용한 구별별 분류율(그림 4에서 신경망 B의 결과)을 프레임 별 인식한 결과를 나타내었다. 표3은 음소별 인식빈도수를 나타낸 것으로 가로는 인식결과이고 세로는 인식대상의 음소이다. 자음의 총 인식률은 88%이고 모음의 인식률은 자음보다 높은 93%이다. 그림 6은 자음뒤에 따라오는 모음을 자음의 종류에 따라 모음을 누적(그림 5의 d)참조)하여 최대값만을 추출하여 백분율로 나

타내었다. 유성자음 (ㄴ, ㄹ, ㄹ) 뒤에 따라오는 모음보다 무성자음 (ㄱ, ㅋ, ㅌ 등) 뒤에 나오는 모음의 누적률이 약간 큼을 알 수 있다. 그림 7은 자음에 대한 누적

률(그림 5의 b참조)을 모두 나타낸 것으로 최대값은 모음보다는 낮으나 대체적으로 60%이상의 값을 가짐을 알 수가 있다.

표 2. 자음 그룹별 분류율

자 음	ㄱ	ㄷ	ㅂ	ㄴ	ㄹ	ㄹ	ㅅ	ㅆ	ㅈ	ㅊ	ㅌ	ㅍ	ㅎ
분류율(%)	92.8	90.5	90.8	95.5	93.5	96.8	84.7	75.6	82.8	85.7	61.8	60.3	74.8
평 균(%)	91.4			95.3			84.4			60.3			74.8
총 평 균	84.3												

표 3. 음소별 인식 빈도수

3-1. 자음 인식

	ㄱ	ㄷ	ㅂ	ㄴ	ㄹ	ㄹ	ㅅ	ㅆ	ㅈ	ㅊ	ㅌ	ㅍ	ㅎ
ㄱ	6	1											
ㄷ		6	1										
ㅂ		1	6										
ㄴ				6	1								
ㄹ					6	1							
ㄹ						7							
ㅅ							5		2				
ㅆ								6	1				
ㅈ									7				
ㅊ										6		1	
ㅌ											1	6	
ㅍ												1	6
ㅎ													7

3-2. 자음 인식

	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ
ㅌ	14								
ㅍ	2	12							
ㅊ		3	11						
ㅌ		1		13					
ㅍ					13				
ㅊ							14		
ㅌ									14

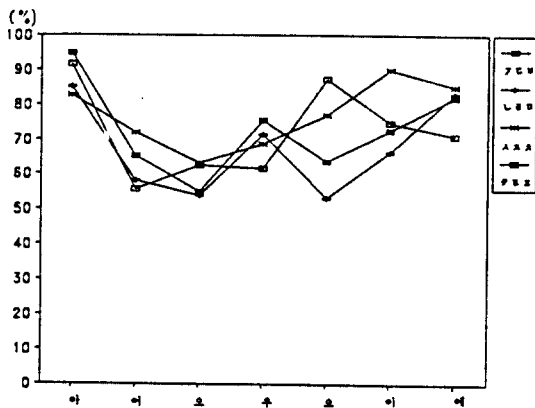


그림 6. 모음의 최대 누적률

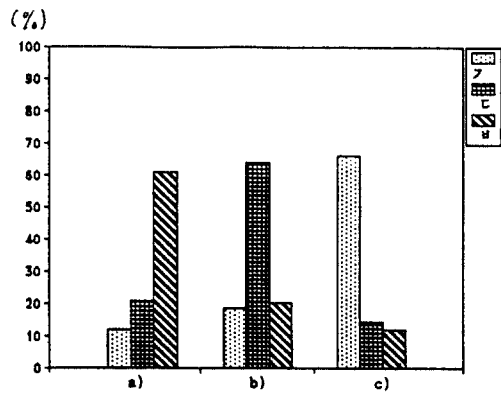


그림 7-1. a) 자음 /ㄱ/에 대한 누적률
b) 자음 /ㄷ/에 대한 누적률
c) 자음 /ㅌ/에 대한 누적률

V. 결 론

본 연구에서는 음성을 음소로 분류하고 나서 각각의 음소를 인식하였다. 자음소의 평균 인식률은 88%이고 모음소의 평균인식률은 93%를 얻었다.

인식에 사용한 신경망은 적은 수의 입력을 대상으로 하기위해서 또한 새로운 어휘가 나타났을때 혼란시 유연성을 주기 위해 소규모로 구성된 여러개의 신경회로망으로 구성하였다.

본 연구가 연속음성에까지를 적용하기 위해서는 각각의 음소가 연음에 의해 변화하는 연구가 필요하다.

참 고 문 헌

1. T.W. Parsons, *Voice And Speech Processing*, McGraw-Hill Book Company, 1986.
2. R.D. Peacock and D.H.Graf, "An Introduction to speech and Speaker Recognition," *Computer*, vol. 23, no. 8, August 1990.
3. L.R. Rabiner and R.W. Schafer, *Digital Processing of speech Signals*, Engelwood Cliffs, N.J., Prentice-Hall, 1978.
4. R.P. Lippmann, "Review of Neural Networks for Speech Recognition," *Readings in Speech Recognition*, Morgan Kufmann, Inc., 1990.
5. H.Sakoe and S.Chiba, "Dynamic Programing Optimazition for Spoken Word Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, pp.43-49, Feb. 1978.
6. Lee, Kai-Fu and H.W. Hon, "Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM," *Proceedings IEEE IASSP*, 123-126, 1988.
7. A.Waibel, T.Hanazawa, G.Hinton, K.Shikano, and K.Lang, "Phoneme Recognition using Time-Delay Neural Network," *IEEE Trans*, Vol. ASSP-37, No. 8, Aug. 1989.
8. A.Waibel, T.Hanazawa, G.Hinton, K.Shikano, and K.Lang : "Phoneme Recognition : Neural Networks vx. Hidden Markov Models," *Proc. IEEE International conference on Acoustics, speech and signal Processing*, New York, NY, April 1988.
9. R.P. Lippman, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, Vol.4, No. 2, pp. 4-22, April 1987.
10. D.J.Burr : "Experiments on Neural Net Recognition

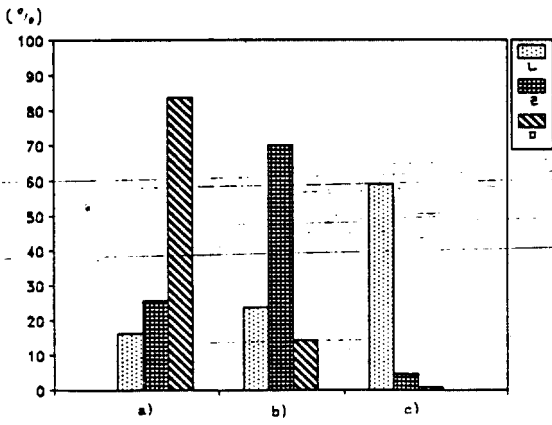


그림 7-2. a) 자음 /m/에 대한 누적률
b) 자음 /b/에 대한 누적률
c) 자음 /n/에 대한 누적률

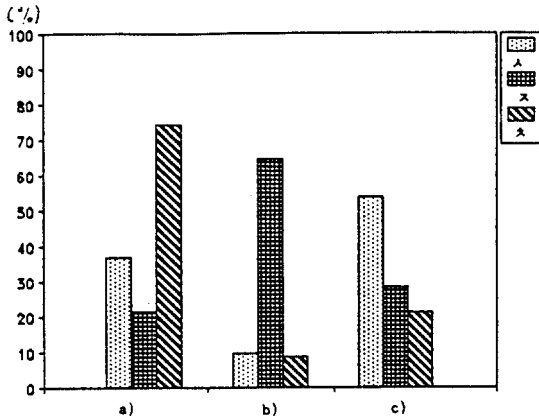


그림 7-3. a) 자음 /k/에 대한 누적률
b) 자음 /g/에 대한 누적률
c) 자음 /ng/에 대한 누적률

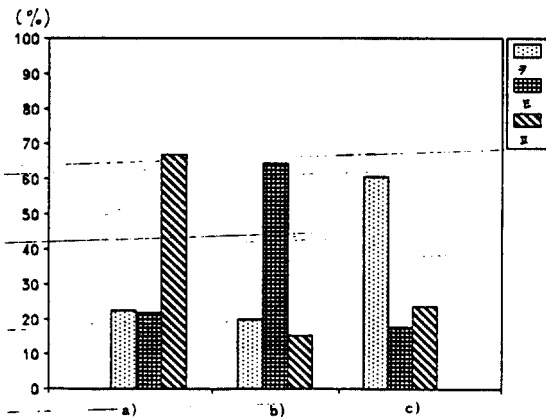


그림 7-4. a) 자음 /p/에 대한 누적률
b) 자음 /t/에 대한 누적률
c) 자음 /k/에 대한 누적률

