

## 신경망을 이용한 음소분할에 관한 연구

正會員 李 光 錫\*\* 正會員 李 宗 振\* 正會員 高 時 暎\*\*\*  
 正會員 許 康 仁\* 正會員 金 明 起\*

### A Study on the Phoneme Segmentation Using Neural Network

Kwang Seok Lee,\*\* Jong Jin Lee,\* Shi Young Gho,\*\*\* Kang In Hur,\*  
 Myong Ki Kim\* *Regular Members*

#### 要 約

본 연구에서는 신경망으로 음성신호를 음소분할 할수 있는 알고리즘을 제시하고 시뮬레이션을 통해 타당성을 검토하였다. 제안된 신경망은 1개의 은닉층을 가지는 다층 인식자로 구성하였다. 실험결과 연속모음과 6개 지명음을 자료로 하여 종속화자인 경우 100%, 독립화자인 경우 99.4%, 그리고 6개 지명음을 동시에 학습하여 각 지명음에 대해 음소분할한 결과는 94.5%의 정합율을 얻었다.

#### ABSTRACT

In this paper, we proposed a method of segmenting speech signal by neural network and its validity is proved by computer simulation. The neural network is composed of multi-layer perceptrons with one hidden layer. The matching accuracies of the proposed algorithm are measured for continuous vowel and place names. The resulting average matching accuracy is 100% for speaker-dependent case, 99.4% for speaker-independent case and 94.5% for each place name when the neural network is trained for 6 place names simultaneously.

#### I. 서 론

현재 세계적으로 실용화 되고 있는 음성인식 장치는 단어나 음절과 같은 단위로 인식하고 있다.<sup>(3),(13)</sup> 1980년 이후 국내에서도 Dynamic time warping (DTW)<sup>(3)</sup>, Vector quantization (VQ)<sup>(14)</sup>, Hidden markov model (HMM)<sup>(6)</sup> 방식등을 도입하여 고품질 단어, 연결 숫자음인식에 대한 연구가 계속되어 왔

다. 실제 단어나 음절로서 인식하는 경우 표준패턴은 조음결합의 영향을 받은 것이되며 표준패턴의 수도 많아서 패턴 매칭시간이 길어지게 된다. 따라서 이를 해결하기 위한 방법으로 음성의 인식단위에 관계없이 표준패턴의 수가 일정한 음소단위로 인식하는것이 중요하며 초성, 중성 및 종성의 음소들로 구성되며 조합성이 뛰어난 한국어 음성의 경우 더욱 필요하다. 그러나 인식단위를 음소로 하면 전 발생시간 중에 조음영향을 받아 변형되는 부분의 비율이 크게 되어 양측의 음소간에서 나타나는 특징의 중간적인 성질이 나타나므로 그대로 인식하는 것은 각 음소간에 서로 존재하지 않는 음소를 존재하고 있는 것처럼 잘못

\* 東亞大學校 電子工學科

Dept. of Electronic Engineering, Dong-A University

\*\* 昌原專門大學 電子通信科

\*\*\* 慶北産業大學 電子工學科

論文番號: 92-48(接受1991. 9. 26)

판단하는 경우가 많게된다. 이러한 문제는 음성신호에서 조음결합의 영향을 받고 있는 부분에 대해 음소를 표현하고 있는 부분과 해당음소에 가까운 조음결합부분을 함께 분할하면 음소단위의 특징이 분할영역내에 나타나므로 음소단위의 분할이 가능하게 된다<sup>(4)</sup>. 음소분할의 연구는 Klatt 등<sup>(1)</sup>이 스펙트로그램을 관찰하여 음소를 분할하였고, Rabiner 등<sup>(2)</sup>은 영교차율과 에너지 그리고 Zue 등<sup>(10)</sup>은 LPC 스펙트럼의 대역 에너지 및 변화도를 이용하여 분할하였다. 본 연구에서는 음소를 표현하고 있는 부분과 해당음소에 가까운 조음결합부분을 함께 분할하는 방법으로 병렬처리기법인 신경망<sup>(5),(7),(11),(12),(13)</sup>을 이용하여 특정화자에 대해 음소단위의 분할에 대한 학습을 수행한 후 중속화자(동일화자), 독립화자(다른화자)의 경우에 대해 자동적으로 음소의 경계를 분할 할 수 있는 알고리즘을 제안하였다. 음소분할을 위한 특징 파라미터로는 켈스트럼계수의 크기인 동적척도(Dynamic measure)값과 에너지를 이용하였다.

## II. 신경망에 의한 음소분할

### 1. 오차후향전파 학습법

다층구조 신경망의 학습에는 오차후향전파학습법이 널리 이용되고 있다. 오차후향전파 학습법은 일반화된 Delta rule로서 목표 출력값과 실제 출력값과의 오차에 해당하는 비용함수(Cost function)의 값을 최소화 시키는 가중치를 찾는 방법으로서 학습방법은 순방향 처리와 역 방향 처리를 하며 전자는 입력 벡터에 대하여 입력층에서 출력층으로의 unit의 상태를 계산하여 가며 후자는 출력층에 접속되어있는 가중치로부터 시작하여 입력층의 방향으로 오차변화에 대한 접속 가중치의 변화를 수정하여 가며 이 두가지 처리로서 오차가 수렴할때 까지 반복수행하여 학습을 진행한다.

그림 1에 일반적인 다층구조 신경망을 보인다.

그림 1에서 입력층에 가까운 순서대로 I층, J층, K층이라고 하고 I층의 unit i로 부터 은닉층인 J층의 unit j로의 접속 가중치를  $w_{ij}$ 로 하고 다음을 가정하면 입력층과 은닉층의 관계는 식(1), (2)와 같이 쓸 수 있다.

$X_i$  : 입력벡터에서 입력층의 unit i로의 입력

$Y_i$  : 입력벡터에서 입력층의 unit i의 상태

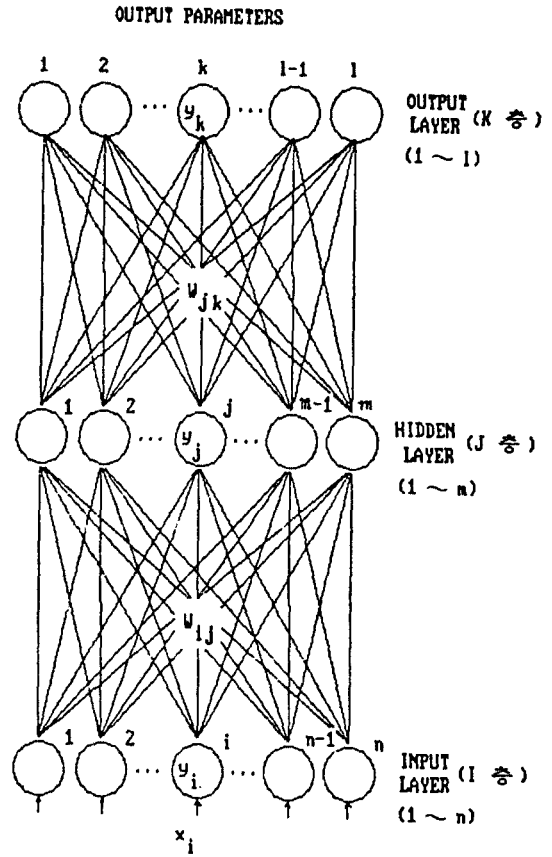


그림. 1 다층 신경망  
Fig. 1 Multi-layer neural network.

$X_i$  : 입력벡터에서 은닉층의 unit j로의 입력

$Y_i$  : 입력벡터에서 은닉층의 unit j의 상태

$W_{ij}$  : 입력층의 unit i에서 은닉층의 unit j로의 접속가중치

$Z_j$  : 입력벡터에서 은닉층의 unit j의 목표출력치

$\Theta_j$  : unit j의 임계치

$$X_j = \sum_i (X_i W_{ij} - \Theta_j) \quad (1)$$

$$Y_j = 1 / (1 + e^{-x_j}) \quad (2)$$

이때, 신경망의 상태의 오차평가는 다음 함수로서 행한다.

$$E = \sum_j (Y_j - Z_j)^2 / 2 \quad (3)$$

가중치  $W_{ij}$ 는  $-\partial E / \partial W_{ij}$ 에 비례하여 변하고 있으며  $W_{ij}$ 의 변화량  $\Delta W_{ij}$ 는 다음식으로 결정된다.

$$\Delta W_{ij} = -\epsilon \frac{\partial E}{\partial W_{ij}} \quad (4)$$

$\partial E / \partial W_{ij}$ 의 값은 식(5)과 같이 계산된다.

$$\frac{\partial E}{\partial W_{ij}} = \sum_k \frac{\partial E}{\partial Y_k} \frac{dY_k}{dX_j} \frac{\partial X_k}{\partial W_{ij}} \quad (5)$$

여기서 식(4), (3), (2)에서 다음식을 각각 얻을 수 있다.

$$\frac{\partial E}{\partial Y_i} = Y_i - Z_i \quad (6)$$

$$\frac{dY_j}{dX_j} = Y_j(1 - Y_j) \quad (7)$$

$$\frac{\partial X_k}{\partial W_{ij}} = Y_i \quad (8)$$

그러므로 식(5)는 다음과 같이 쓸 수 있다.

$$\frac{\partial E}{\partial W_{ij}} = \sum_k (Y_k - Z_k) Y_i (1 - Y_j) Y_k \quad (9)$$

여기서 가중치의 초기치는 학습을 위해 랜덤하게 설정되어지며 입력벡터 1개에 대하여 1개의 출력벡터가 결정되어 전 가중치에 대하여  $\partial E / \partial W_{ij}$ 가 계산되어진다. 그리고 입력벡터가 전부 주어진 후 가중치의 변화량  $\Delta W_{ij}$ 가 구해지고 unit 간의 접속가중치는 학습횟수에 따라 조정되어진다. 여기서 J층의 unit j의 동작이 오차로서 영향을 미치는 것은 다음층 K의 unit k를 통할 때 뿐이라면 식(5)의  $\partial E / \partial Y_j$ 를 다음식으로 나타낼 수 있다.

$$\begin{aligned} \frac{\partial E}{\partial Y_j} &= \sum_k \frac{\partial E}{\partial Y_k} \frac{dY_k}{dY_j} \frac{\partial X_k}{\partial Y_j} \\ &= \sum_k \frac{\partial E}{\partial Y_k} Y_k (1 - Y_k) W_{jk} \end{aligned} \quad (10)$$

여기서  $W_{jk}$ 는 은닉층의 unit j에서 출력층의 unit

k로의 접속가중치이다. 이것으로부터 K층의 전체 unit에 대하여  $\partial E / \partial Y_k$ 를 안다면 J층의 unit에 대해서도 쉽게 구해지므로 출력층에서 입력층까지의 학습이 이루어진다.

여기서 가중치변화량  $\Delta W$ 는 일반적으로 오차의 진동을 감소시켜 수렴을 빨리 할 수 있도록 다음식으로 결정한다.

$$\Delta W_{ij}(t+1) = -\epsilon \frac{\partial E}{\partial W_{ij}} + \alpha \Delta W_{ij}(t) \quad (11)$$

여기서 t는 학습횟수이며, 파라미터  $\epsilon$ ,  $\alpha$  각각 학습율, 모멘텀을 나타내며 학습능력을 변화시키는 임의의 값이다.

## 2. 음소분할 신경망의 구성

음성의 자동 음소분할을 위한 신경망은 1개의 은닉층이 존재하는 다층인식자(multi-layer perceptron)로 구성하였으며 그림 2에 보인다. 그림 2에서 각 unit는 시그모이드(Sigmoid) 함수를 사용하였으며 학습효율을 높이기 위해서 입력층과 은닉층에 상태가 항상 1이 되는 unit를 각각 1개씩 둔다. 이렇게 하면 은닉층과 출력층 unit의 임계치는 항상 1을 취하는 unit에서의 입력으로 생각할 수 있어 임계치의 학습은 상태가 항상 1인 unit에서의 접속 가중치 학습으로 처리될 수 있다. 입력층에는 일정한 수(21개)의 파라미터를 인가하고 1프레임씩 증가시켜가는 방법을 취했으며 입력 unit를 21개로 설정한 이유는 음소를 표현하고 있는 부분과 각 음소에 가까운 조음결합 부분을 동시에 분할할 경우 210 ms(21프레임)이하에 경계가 1개 이상 있기 때문이다. 음성의 전 프레임수와 같은 입력 파라미터가 주어지며 입력 파라미터가 존재하지 않는 시작 또는 끝부분에는 0을 부여하였다. 그리고 출력 unit는 입력층의 중심인 11번째 unit에 주어지는 파라미터를 갖는 프레임이 경계위치인가를 판단하여 결정한다.

그림 2에서 입력은 입력벡터의 수가 n개이면 n번 입력이 들어가므로 m번째 j층의 입력은 다음과 같다.

$$\begin{aligned} X_{jm} &= \sum_i X_{im} W_{ij} \\ &= \sum_i X_{im} W_{ij} + X_{0i} W_{0j} \\ &= \sum_i X_{im} W_{ij} - W_{0i}, \quad m=1,2,\dots,n \end{aligned} \quad (12)$$

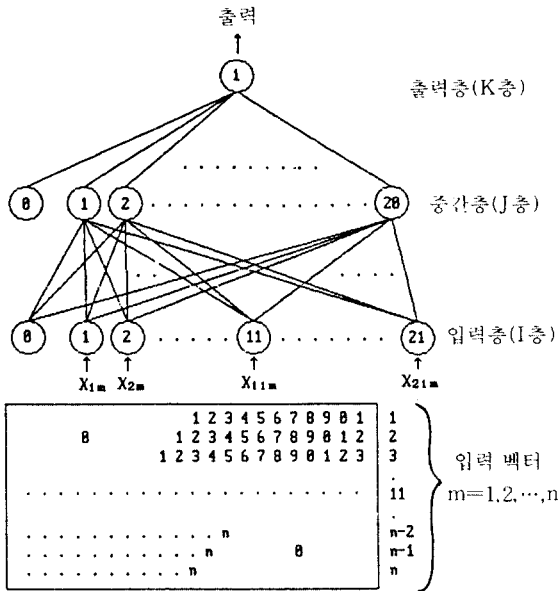


그림. 2 음소분할 신경망의 구성  
Fig. 2 The configuration of neural network for the phoneme segmentation.

식(12)는 식(1)의  $\Theta_j$ 를  $W_{0j}$ 로 한 것과 같다. 그러므로 상태 -1인 unit에서의 가중치  $W_{0j}$ 는 unit j의 임계값으로 생각할 수 있다.  
그러므로 그림 2의 각 층의 unit의 상태는 다음과 같다.

입력층 : unit 1~unit 21, unit 0  
 $X_{im}$  : m번째 입력벡터의 unit i로의 입력  
 $Y_{im}$  : unit i의 상태  
 $Y_{0m} = X_{im}$   
 $Y_{0m} = -1$   
 여기서  $i=0,1,\dots,21$   
 $m=1,2,\dots,n$

은닉층 : unit 1~unit 20, unit 0  
 $X_{jm}$  : m번째 입력벡터의 unit j로의 입력  
 $Y_{jm}$  : 입력  $X_{jm}$ 를 얻었을 때의 unit j의 상태  
 $W_{ij}$  : 입력층의 unit i에서 은닉층의 unit j로의 접속가중치

$$X_{jm} = \sum_i X_{im} W_{ij}$$

$$Y_{jm} = \frac{1}{1 + e^{-X_{jm}}}$$

$$Y_{0m} = -1$$

여기서  $i=0,1,\dots,21$   
 $j=1,2,\dots,20$   
 $m=1,2,\dots,n$

출력층 : 출력 unit k (1개)  
 $X_{km}$  : 출력 unit k로의 m번째 입력  
 $Y_{km}$  : 입력  $X_{km}$ 을 얻었을 때의 출력 unit k의 상태  
 $W_{jk}$  : 은닉층의 unit j에서 출력 unit k로의 접속가중치  
 $X_{km} = \sum_j Y_{jm} W_{jk}$   
 $Y_{km} = \frac{1}{1 + e^{-X_{km}}}$   
 여기서  $j=0,1,\dots,20$   
 $m=1,2,\dots,n$

### 3. 학습

#### 1) 학습데이터

신경망의 입력 파라미터는 차수의 항들이 각각 연결성을 유지하며 변화하는 켈스트럼계수의 변화에 대한 크기인 동적척도값을 사용하였다. i번째 프레임의 j차 켈스트럼 계수를  $C_{ij}$  ( $j=1\sim 16$ ), 에너지를  $C_{i0}$ 로 하면 i번째 프레임의 j차 켈스트럼 계수열의 회귀 직선의 기울기  $a_{ij}$ 는 다음과 같다.

$$a_{ij} = \frac{\sum_k C_{i+k,j} k W_k}{\sum_k k^2 W_k}, \quad j=0\sim 16, -L \leq k \leq L \quad (13)$$

여기서,  $W_k$ 는 폭이  $2L+1$ 인 창함수이다. i번째 프레임의 동적척도  $D_i$ 는 다음과 같이 정의된다.

$$D_i = a_{i0}^2 / 2 + \sum_j a_{ij}^2 \quad (14)$$

실험에 사용되는 학습데이터는 원만한 학습을 위해 기준경계를 중심으로 그림 3과 같이 1.0, 0.8, 0.3, 0으로 설정 하였으며 지명음/부산(Pusan)의 동적척도값과 학습데이터의 예를 그림 4에 나타내었다.

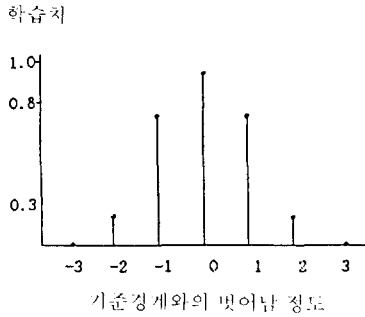


그림. 3 학습 데이터  
Fig. 3 Learning data.

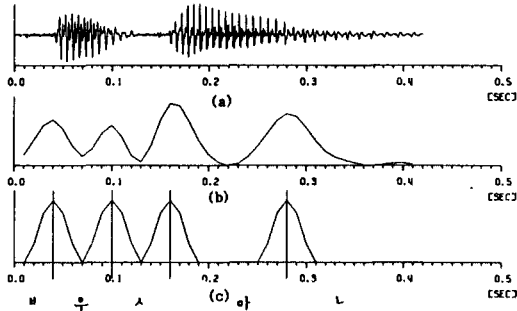


그림. 4 지명음/부산/의 동적척도값과 학습데이터  
(a) 음성파형 (b) 동적척도값 (c) 학습데이터  
Fig. 4 The example of place name /Pusan/.  
(a) Speech wave (b) Dynamic measure  
(c) Learning data

여기서, 창 의 폭은 음소분할에 필요한 정보를 찾을 수 있는 범위를 정하는 값으로 실험적으로 L=5로 하였다.

2)가중치 변화량

학습의 수렴속도를 빠르게 하기위해서 1회의 가중치 변화의 절대값에 제한을 설정하고 1회의 변화량이 너무 크지 않도록 학습의 진동을 억제하면서 오차 E를 최소화로 하여 가중치 변화량 ΔW를 계산한다. 그림 2에서 신경망의 상태의 오차평가는 m개의 입력에 대해 식(3)의 형태로 주어진다.

먼저 순방향 처리에서 은닉층에서 출력층으로의 가중치는 다음 순서로 구한다.

ΔW<sub>jk</sub>(t) : t번째 학습에서의 가중치 W<sub>jk</sub>의 변화량, (j=0,1,...,20)

$$\frac{\partial E}{\partial W_{jk}} = \sum_m \frac{\partial E}{\partial Y_{km}} \frac{dY_{km}}{dX_{km}} \frac{\partial X_{km}}{\partial W_{jk}} = \sum_m \{ (Y_{km} - Z_{km}) Y_{km} (1 - Y_{km}) Y_{jm} \}, m=1,2,\dots,n \quad (15)$$

$$\Delta W_{jk}(t) = -\epsilon \frac{\partial E}{\partial W_{jk}} + \alpha \Delta W_{jk}(t-1) \quad (16)$$

$$\begin{aligned} \text{여기서, } \Delta W_{jk}(t) &= -\eta, & \Delta W_{jk}(t) &\leq -\eta \\ \Delta W_{jk}(t) &= \Delta W_{jk}(t), & -\eta &< \Delta W_{jk}(t) < \eta \\ \Delta W_{jk}(t) &= \eta, & \Delta W_{jk}(t) &\geq \eta \end{aligned}$$

그리고 역방향 처리에서 은닉층에서 입력층으로의 가중치는 식(5), (10)에서 다음 순서로 구한다.

ΔW<sub>ij</sub>(t) : t번째 학습에서의 가중치 W<sub>ij</sub>의 변화량, i=0,1,...,21, j=1,...,20

$$\frac{\partial E}{\partial W_{ij}} = \sum_m \frac{\partial E}{\partial Y_{jm}} \frac{dY_{jm}}{dX_{jm}} \frac{\partial X_{jm}}{\partial W_{ij}} = \sum_m \{ (Y_{km} - Z_{km}) Y_{km} (1 - Y_{km}) W_{jk} Y_{jm} (1 - Y_{jm}) Y_{im} \}, m=1,2,\dots,n \quad (17)$$

이것 식(15)를 이용하면 식(18)이 된다.

$$\frac{\partial E}{\partial W_{ij}} = \sum_m \{ \frac{\partial E}{\partial W_{jk}} W_{jk} (1 - Y_{jm}) Y_{im} \}, m=1,2,\dots,n \quad (18)$$

$$\Delta W_{ij}(t) = -\epsilon \frac{\partial E}{\partial W_{ij}} + \alpha \Delta W_{ij}(t-1) \quad (19)$$

이 된다.

$$\begin{aligned} \text{여기서, } \Delta W_{ij}(t) &= -\eta, & \Delta W_{ij}(t) &\leq -\eta \\ \Delta W_{ij}(t) &= \Delta W_{ij}(t), & -\eta &< \Delta W_{ij}(t) < \eta \\ \Delta W_{ij}(t) &= \eta, & \Delta W_{ij}(t) &\geq \eta \end{aligned}$$

파라미터 ε, α, η는 각각 학습율, 모멘텀항 및 학습의 임계치를 나타내며 이는 학습능력을 변화시키므로

신경망의 성능을 향상시키기 위해 적절한 값으로 설정하여야 한다.

4. 알고리즘의 흐름도

신경망에 의한 음성의 음소분할 흐름도는 그림 5와 같다.

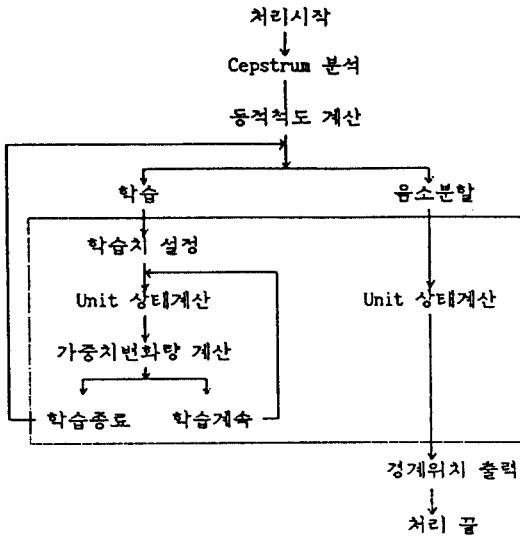


그림. 5 알고리즘의 흐름도  
Fig. 5 The flow chart of algorithm.

Ⅲ. 실험결과 및 고찰

실험에 사용된 음성자료는 10명의 화자가 5번씩 발성한 /아이우에오/, /서울/, /부산/, /대구/, /광주/, /인천/, /대전/이다. 분석조건은 5 KHz의 LPF를 통과시켜 10 KHz 샘플링, 12 bit로 양자화한 데이터에 1차 차분으로 고역강조를 행한 후 20 ms의 hanning 창, 10 ms의 프레임 주기로 설정하였다. 학습성취도는 초기 가중치의 설정상태와 학습대상에 따라 크게 좌우되므로 여러 경우의 정합율을 검토하여 은닉층에서 출력층, 입력층에서 은닉층으로의 연결 가중치는 각각  $-3 < W_{jk} < 3$ ,  $-5 < W_{ij} < 5$  사이의 난수로 설정 하였으며 학습률과 모멘텀 및 학습의 임계치는 각각 1.0, 0.9, 0.01로 하였다. 은닉층의 갯수는 20개로 설정하였고 학습회수는 500회 이상에서 변화가 거의 없으므로 500회로 하였다. 음소분할에 대한

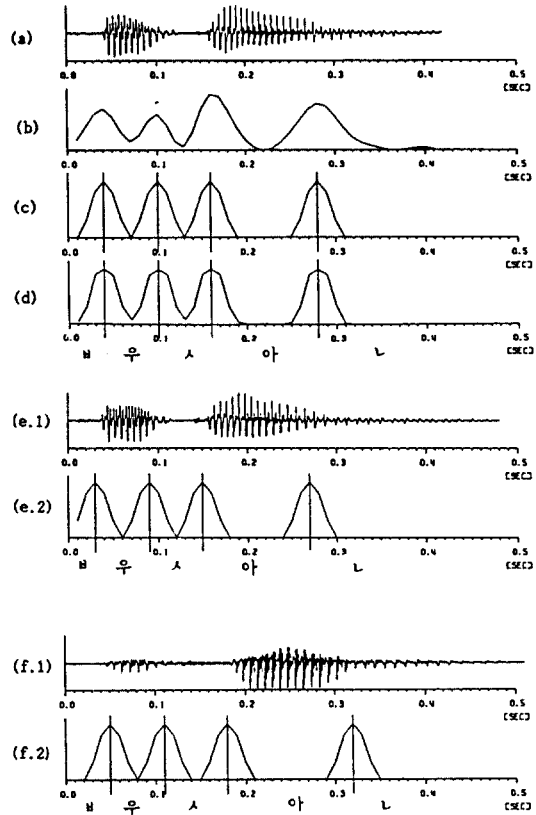


그림. 6 음소분할의 결과 /부산/  
(a) 학습 음성, (b) 동적 척도 값, (c) 학습 데이터  
(d) 500회 학습후의 학습데이터 상태  
(e.1) 종속 화자의 실험 음성  
(e.2) (e.1)의 음소분할 결과  
(f.1) 독립 화자의 실험 음성  
(f.2) (f.1)의 음소분할 결과  
Fig. 6 Results of phoneme segmentation /Pusan/.  
(a) Learning speech, (b) Dynamic measure  
(c) Learning data,  
(d) Status of learning data after 500-times learning  
(e.1) Test speech of speaker-dependent  
(e.2) The result of phoneme segmentation of (e.1)  
(f.1) Test speech of speaker-independent  
(f.2) The result of phoneme segmentation of (f.1)

정합율의 평가는 다음과 같이 정의 한다.

$$\text{정합율} = ((A+B \times 0.5 + C \times 0. - D) / M) \times 100(\%) \quad (21)$$

- 여기서, A : 기준경계와 일치 및 2 프레임 오차.
- B : 기준경계와 3 프레임 오차.
- C : 기준경계와 4 프레임 오차.
- D : 기준경계와 5 프레임 이상의 오차.
- M : 음소분할 대상 자료의 수.

지명음 /부산(Pusan) /에 대한 음소분할의 결과를 그림 6에 보였다. 그림 6(a)는 학습하기위한 음성, (b)는 동적척도의 값, (c)는 학습데이터, (d)는 500회 학습후의 데이터, (e.1)는 종속화자의 실험음성, (e.2)는 학습된 신경망으로 (e.1)을 음소분할한 결과, (f.1)은 독립화자의 실험음성, (f.2)는 학습된 신경망으로 (f.1)을 음소분할한 결과를 보인다. 그림 6의 (d)에서 학습음성에 대한 음소분할의 경계는 4, 10,16,28번째 프레임이며, (e.2)의 종속화자는 3,9, 15,27번째 프레임 그리고 (f.2)의 독립화자는 5,11,

18,32번째 프레임으로 나타났으며 음소분할의 결과가 양호함을 보인다.

그리고 독립화자에 대해 표 1은 각 음성을 각각 학습시킨 후의 각음성에 대해 음소분할한 정합율의 결과를 보였으며 표 2는 6개 지명음을 동시에 학습시킨 후 각 지명음에 대한 정합율을 보였다.

종속화자에서 연속모음, 지명음들의 정합율은 10명이 5번 발생한 것 중 하나는 학습을 위한 음성으로 사용되므로 실험음성은 나머지 4번 발생한 것이므로 식 (21)에서 M=40이 되며 정합율은 100%로 나타났다. 독립화자인 경우 정합율은 학습을 위한 음성을 선택한 후 종속화자에서 사용한 실험음성에 대해 (M=40) 99.4%로 나타났다. 동시 학습인 경우 10명이 5번 발생한 실험음성(M=50)에 대해 94.5%로 나타났다. 그리고 독립화자인 경우 연속모음에서 최대 2 프레임, 지명음에서 최대 3프레임의 오차가 발생하였다. 6개 지명음을 동시학습시켜 각 지명음에 대한 오차는 최대 6프레임의 오차가 발생하였다. 이것은 6개 지명음을 동시에 발생한 것을 사용하지 않고 각

표. 1 정합율의 결과(독립화자)  
Table. 1 The result of matching rate(speaker-independent).

평균 정합율 음성원	음소경계 1	음소경계 2	음소경계 3	음소경계 4	총평균 정합율
아이우에오	아/이	이/우	우/에	에/오	100%
	100	100	100	100	
서 울	ㅅ/어	어/우	우/르		95.8%
	100	87.5	100		
부 산	ㅂ/우	우/ㅅ	ㅅ/아	아/ㄴ	100%
	100	100	100	100	
대 구	ㄷ/애	애/ㄱ	ㄱ/우		100%
	100	100	100		
광 주	ㄱ/와	와/ㅇ	ㅇ/ㅅ	ㅅ/우	100%
	100	100	100	100	
인 천	이/ㄴ	ㄴ/ㅅ	ㅅ/어	어/ㄴ	100%
	100	100	100	100	
대 전	ㄷ/애	애/ㅅ	ㅅ/어	어/ㄴ	100%
	100	100	100	100	
					99.4%

표. 2 정합율의 결과(6개 지명음의 동시 학습)  
Table. 2 The result of matching rate.

평균 정합율		음소경계 1	음소경계 2	음소경계 3	음소경계 4	총평균 정합율
음성원	서 울	ㅅ/어	어/우	우/ㄹ		
		100	33.0	100.0		77.7%
부 산		ㅂ/우	우/ㅅ	ㅅ/아	아/ㄹ	
		100	100	100	100	100%
대 구		ㄷ/애	애/ㄱ	ㄱ/우		
		100	100	100		100%
광 주		ㄱ/와	와/ㅇ	ㅇ/ㅅ	ㅅ/우	
		100	100	100	100	100%
인 천		이/ㄹ	ㄹ/ㅅ	ㅅ/어	어/ㄹ	
		56.0	100	100	100	89.0%
대 선		ㄷ/애	애/ㅅ	ㅅ/어	어/ㄹ	
		100	100	100	100	100%
						94.5%

지명음을 합쳐 6개 지명음으로 편집하였으므로 조음 결합의 영향으로 인한 캡스트럼계수의 변화가 불분명하여진 것으로 사료된다.

### V. 결 론

본 연구에서는 음소를 표현하고 있는 부분과 해당 음소에 가까운 조음결합부분을 함께 분할하는 방법으로서 병렬처리기법인 신경망을 이용하여 특정화자에 대해 음소단위의 분할에 대한 학습을 수행한 후 종속화자(동일화자), 독립화자(다른화자)의 경우에 대해 자동적으로 음소의 경계를 분할 할 수 있는 알고리즘을 제안하였다. 제안된 신경망의 입력으로는 캡스트럼계수의 크기인 동적최도값과 에너지를 사용하였으며 입력층에는 입력데이터를 1프레임씩 밀어 가는 방법을 취했기 때문에 음성원의 길이에 대한 제한이 필요하지 않으며 은닉층의 unit는 20개, 출력층은 1개로 구성되었다. 실험결과, 종속화자와 독립화자의 경우 음소 분할의 타당성을 인정하였고 6개 지명음의 학습에서는 6개 지명음을 동시에 발생한 것을 사용하지 않고 각 지명음을 합쳐 6개 지명음으로 편집하였으므로 조음결합의 영향이 나타난것으로 생각된다. 그러므로 제안된 신경망에 조음결합을 고려한 연속음성으로 학습을 하면 더욱 효율적이라 생각되

며 사용된 입력과 음성분석에서 구할 수 있는 특징 파라미터들의 적합한 적용을 고려하면 더욱 좋은 정합율을 얻을 수 있을것으로 사료되며 향 후 제시된 신경망을 수정 구성하면 음소단위의 음성인식에 응용할 수 있을 것으로 기대된다.

### 참 고 문 헌

1. D.H. Klatt, K.N. Stevens, "Strategies for Recognition of Spoken Sentences from Visual Examination of Spectrogram," BBN, Inc., Cambridge, Mass., BBN Rep., 2154, June 1971.
2. L.R. Rabiner, M.R. Sambur, "Some Preliminary Experiments in the Recognition of Connected Digits," IEEE, ASSP-24, No.2, Apr. 1976.
3. 김순협, "한국어 음성의 분석과 자동인식에 관한 연구," 연세대학교 대학원, 1982
4. 板橋 秀一, 西野 博二, 信吾, "音聲의 세그멘테이션における 重み窓의 效果," 電子情報通信學會 論文集, vol. J68-D, No.3, pp.410-411, 1985.
5. DE.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning representation by back-



propagation errors," NATURE, pp.533-536, 1986.

6. L.R. Rabiner and B.H.Juang, "An introduction to hidden Markov models," IEEE ASSP Magazine, vol.3, No.1, pp.4-16, 1986.

7. R.P. Lippmann, "An introduction to computing with Neural Nets," IEEE ASSP Magazine, 4-22, April, 1987.

8. L.R.Rabiner, "A Linear predictive front end processor for speech recognition in noise environments," Proc. ICASSP, pp.1324-1327, 1987.

9. Y. Tohkura, "An introduction to computing with Neural Nets," IEEE ASSP Magazine, 4-22, April, 1987.

10. J.R. Glass, V.W. Zue, Multi-Level Acoustic

Segmentation of Contineous Speech, ICASSP 88, pp.429-432

11. 金寺登, "ニューラル・ネットを用いた連続音聲の音韻セグメンテーション," 音響學會講演論文集, 2-P-6, pp.213-214, 1988.

12. 河原 英紀, "神経回路網モデルと音聲認識," 人工知能學會誌, vol.3, No.4, pp.453-460, 1988.

13. Matsucka, "Syllable Recognition using integrated Neural Networks," IEICE, vol. SP 87-101, Dec, 1989.

14. 최계관, "음성처리및 인식의 기초연구," KOSEF 86-0203-04, 1989.

15. 이광석, 이종진, 고시영, 허강인, 김명기, "신경망을 이용한 음소구분화에 관한 연구," 계통종합학술발표회 논문집, vol.10, No.2, pp.187-189, 1991.



李 光 錫(Kwang Seok Lee) 正會員  
1957年 9月 7日生  
1983年 2月: 東亞大學校 電子工學  
科 卒業(工學士)  
1985年 2月: 東亞大學校 大學院 電  
子工學科 卒業(工學  
碩士)  
1992年 2月: 東亞大學校 大學院  
(工學博士)

1986年 3月~現在: 昌原專門大學 電子通信科 助教授



李 宗 振(Jong Jin Lee) 正會員  
1958年 8月 20日生  
1982年 2月: 東亞大學校 電子工學  
科 卒業(工學士)  
1986年 2月: 東亞大學校 大學院 電  
子工學科 卒業(工學  
碩士)  
1990年 3月~現在: 長安專門大學 電  
算科 專任講師(東亞  
大學校大學院 電子工  
學科 博士課程)



高 時 暎(Shi Young Gho) 正會員  
1952年 8月 16日生  
1979年 2月: 嶺南大學校 電子工學  
科 卒業(工學士)  
1983年 2月: 嶺南大學校 大學院 電  
子工學科 卒業(工學  
碩士)  
1990年 2月: 東亞大學校 大學院 電  
子工學科(博士課程修  
了)

1980年 3月~現在: 慶北產業大學 電子工學科 教授



許 康 仁(Kang In Hur) 正會員  
1955年 2月 20日生  
1980年 2月: 東亞大學校 電子工學  
科 卒業(工學士)  
1982年 2月: 東亞大學校 大學院 電  
子工學科 卒業(工學



金 明 起(Myong Ki Kim) 正會員

1930年 2月 19日生

1951年 8月 : 海軍士官學校 卒業  
(理學士)

1958年 6月 : 美國海軍工科大学 電  
子工學科 卒業(工學  
士)

1966年 2月 : 서울대학교 大學院 電  
子工學科 卒業(工學  
碩士)

1976年 2月 : 東亞대학교 大學院(工學博士)

1977年 3月 ~ 現在 : 東亞대학교 工科大学 電子工學科教授