신경망을 이용한 연속 숫자음 인식에 관한 연구

A Study On Continuous Digits Recognition Using the Neural Network

이 성 권*, 김 순 협*
(Seong Kwon Lee*, Soon Hyob Kim*)

※본 논문은 한국 과학재단(과제번호 95-0100-22-01-3)의 연구 지원에 의해 연구된 것입니다.

요 약

본 논문은 음성 다이얼링 시스템을 구현하기 위한 한국어 단독 숫자음 및 연속 숫자음 인식에 관한 것이다. 단독 숫자음의 인식은 미지의 입력 음성을 재귀 신경망을 이용하여 모델링된 각 모델에 인가하고, 신경 회로망의 출력 노드의 상태열을 검사하여 적절한 상태 전이를 하며 최고의 확률값을 출력하는 모델을 인식된 결과로 출력한다. 연속 숫자음의 인식은 미지의 연속 숫자음을 재귀 신경 회로망을 이용한 연속 숫자음 모델에 입력하고, 신경 회로망의 출력에 대하여 적절한 상태 전이에 대한 검사와 레벨 빌딩(Level Building)을 수행하여 최소의 오차를 가지는 모델열을 인식된 결과로 출력한다. 재귀 신경 회로망을 이용하여 음절 모델을 만드는 과정에서 재귀 노드는 예상치가 주어지지 않으므로 신경 회로망의 학습에서 제외되어 현저한 학습 속도의 저하를 가져온다. 따라서 본 논문에서는 재귀 신경 회로망의 학습 속도를 향상시키기 위한 2가지 방법을 제안 한다. 첫번째는 재귀 신경 회로망의 재귀 노드의 예상치를 실험적으로 주어줌으로써 학습 속도의 향상을 도모하였다. 두번 째는 음절 모델의 출력노드의 개수와 음절 모델의 세그먼트 경계를 알고리듬을 이용하여 자동적으로 조절하였다. 실험결과, 단독어의 경우 음절 '에'를 포함하는 한국어 11개의 숫자음에 대하여 화자 종속의 경우 97.3%, 화자 독립의 경우 80.5%의 인식률을 얻었으며, 연속 숫자음의 경우는 21 종류의 연속 숫자음에 대하여 화자 종속에서 88.2%, 화자 독립의 경우 81.3%의 인식률을 얻을 수 있었다.

ABSTRACT

This paper is a study on the Korean isolated digits and continuous digits recognition for the implementation of voice dialing system. In the Korean isolated digits recognition, after the recognition system inserts unknown isolated digit into the Korean digit models using recurrent neural network and calculates outputs and then output states of Korean digit models are tested. The Korean digit model which have proper output state transition is compared with the others. The recognition result is the Korean digit model index which has proper output state transition and maximum probability. In the Korean continuous digits recognition, proposed recognition system inserts unknown continuous digit inputs into continuous digit models and calculates outputs. The system performs level building and checks proper state transition check, it comes out that the index sequence of continuous digit models which has proper state transition. Finally sequences of continuous digit model index which have minimum accumulated distance are considered as the recognized result. In the procedure of making syllable models, because conventional recurrent neural network has no desired output of recurrent node, the speed of training goes slow. Therefore, this paper proposes two training speed up methods of recurrent neural network. The one is giving the expected value of recurrent node experimentally and the other is adjustment the number of output node and the segment boundary of syllable model automatically with algorithm. Adjustment technique plays the role of finding the minimum number of digit model's segment and re-estimating segment boundary. The experiment was performed for 11 Korean isolated digits including Korean syllable /e/ and the recognition rate is 97.3% in the speaker dependent, 80.5% in the speaker independent. In the continuous digit recognition, the recognition rate is 88.2% in the speaker dependent and 81.3% in the speaker independent for the 21 sorts of continuous digit.

*광운대학교 컴퓨터 공학과
접수일자 : 1998년 1월 14일

Ⅰ. 서 론

현재에는 인간의 두뇌의 신경세포를 모델링하여 인간의 뇌가 수행하는 역할을 하도록 하는 인공 신경 회로망(Artificial Neural Network)에[3] 대한 연구가 활발히 진행되고 있다. 인공 신경 회로망(Artificial Neural Network)은 생물의 신경세포의 동작과 신경 세포간의 상호작용을 모델링한 것으로 1943년 McCulloch와 Pitts[13]에 의해 최초로 제안되었다. 그들의 연구는 정보처리의 관점에서 신경 회로의 동작을 수리적으로 접근한 효시라고 할 수 있다. 그들은 논리 함수를 실현하고자 "1 또는 0"의 출력을 가지는 선형 2진 소자의 동작 특성을 갖는 뉴런을 모델링하였으며, 그 후 1949년에 이르러 Donald O. Hebb[14]는 "뉴런 A와 뉴런 B의 연결 강도는 하나의 뉴런의 활성에 공헌한다면 증가되어야 한다."는 신경 회로망의 학습 방법을 제시하였다. 또한 1958년에 Rossenblatt는 퍼셉트론(Perceptron)[15]을 발표하여 큰 호응을 얻었으나, 1969년에 이르러 Minsky와 Papert[16]가 "퍼셉트론은 XOR문제를 해결할 수 없다."는 것을 증명함으로써 신경 회로망에 대한 연구는 한때 주춤하였다. 그 후 여러 개의 은닉층을 가진 다층 퍼셉트론(Multi-Layer Perceptron)이 제시되어 XOR문제를 해결함으로써 다시 신경 회로망에 대한 연구가 활발하게 진행되어 오고 있다. 재귀 신경 회로망은 1986년에 Jordan이 제시한 재귀 신경 회로망 모델이 주목 받기 시작하면서부터 본격적인 연구가 이루어지기 시작하였다[21]. Jordan이 제시한 재귀 신경 회로망은 재귀 부분이 다음 프레임에 대한 예측 기능과 시변화 특성을 흡수하려는 의도에서 시작되었다. 하지만 Jordan이 제시한 재귀 신경 회로망은 출력층의 모든 노드가 입력층으로 완전 결합(Full Connection)이 이루어지지 않고 일부 노드만이 입력층으로 연결되어 있으며, 시간 지연과 상대 전이를 특정 노드에서만 이루어지도록 제안하였다. Jordan이 제시한 모델을 효시로 하여 많은 재귀 신경 회로망이 제안되고 있다. 현재 주목을 받고 있는 재귀 신경 회로망은 OCON(One Class One Network)구조[22]로 이루어져 있다. 음성 인식을 위한 많은 연구의 목표는 잡음이 있는 환경에서 불특정 화자가 자연스럽게 발음한 대용량 어휘의 연속 음성을 실시간에 기계가 인식하고 이해하는 것이다. 이 목표를 이루기 위하여 연구되어야 할 분야가 많지만 그 중에서 가장 중요한 부분 중의 하나가 연속 음성의 인식이다.[10] 그러나 자연스럽게 발음한 무제한적인 어휘 및 주제의 대화체 음성을 인식하는 시스템은[9] 아직까지 구현이 어렵기 때문에 현재에는 제한된 범위의 어휘에서 연속 음성을 인식하는 음성 인식 시스템을 구현하고 있다.[11][12] 따라서 본 논문에서는 인간의 자연스러운 욕구인 '음성을 통한 인간과 컴퓨터 사이의 인터페이스'에 대한 연구의 일환으로[7], 제한된 범위의 연속 음성 인식 시스템인 '음성 다이얼링 시스템'의 구현에 있어서 필수적인 부분인 연속 숫자음 인식에 그 목적을 두고 있다. 기존의 연결 및 연속 음성을 인식하는 신경 회로망은 대부분 음소 모델을 이용하기 때문에 연결 및 연속 음성에서 음소를 추출하는 음소 세그멘테이션(Segmentation) 과정이 필수적이다. 하지만 연결 및 연속 음성에서의 음소 세그멘테이션은 음소간의 경계가 모호하고 그 경계가 시간에 따라 변하기 때문에 정확한 음소 세그멘테이션이 어렵다.[21] 또한 음소 간의 경계에서 신경 회로망이 정확한 음소를 출력하지 못하는 경우가 많기 때문에 인식률의 저하를 가져온다. 인식 단위를 음소로 하였을 때 따르는 문제점을 피할 수 있는 방법은 인식 단위를 단어 단위로 선정하는 것이다.[5][6] 하지만 이 경우는 모호한 음소 경계에 대한 문제를 해결할 수는 있지만, 반면에 상대적으로 많은 수의 단어 모델을 가져야 한다는 단점을 가진다. 따라서 본 논문에서는 인식 단위를 음절로 하여 음소 간의 경계가 모호한 경우를 해결하였으며[4], 인식 단위를 단어로 선정하였을 경우 보다 훨씬 적은 수의 모델만으로도 연결 및 연속 음성을 인식할 수 있는 장점을 가진다. 또한 재귀 신경 회로망을 사용하여 음절을 모델링할 때 재귀 노드의 예상치가 없으므로 인하여, 학습에서 제외되며 이로 인하여 학습 속도의 지하를 가져온다. 그러므로 본 논문에서는 재귀 노드의 예상치를 실험적으로 주어 줌으로서 학습 속도의 향상을 도모하였다. 그리고 인식 단위를 음절로 선정하였기 때문에 최적의 음절 열을 찾기 위하여 재귀 신경 회로망의 출력에 레벨 빌딩(Level Building) 알고리듬을 적용하였다.[1][8]

Ⅱ. 신경 회로망을 이용한 음성 인식

2.1 재귀 신경 회로망

신경 회로망을 이용한 어플리케이션(Application)을 살펴보면, 분류하고자 하는 패턴으로부터 정적인 특징을 추출하여 사용하는 경우와 분류하고자 하는 패턴의 정적인 특징 및 동적인 특징을 함께 사용하는 경우로 나누어 볼 수 있다. 첫번째 방식으로 음성 인식에 접근하는 경우는 사전에 사람에 의한 세그멘테이션 과정이 필수적으로 수반되어야 하며, 세그멘테이션 경계의 오류가 발생할 수 있기 때문에 인식률의 저하를 가져온다. 또한 음성은 정적인 패턴이 아니라 시간에 따라, 화자에 따라, 발성 시간 및 장소에 따라 변화하는 동적인 패턴이므로 정확한 세그멘테이션 경계를 찾기 어렵다. 두번째의 경우는 입력 프레임을 중첩시켜 신경 회로망에 인가하거나, 출력층에 재귀 노드를 두어 음성의 정적인 특징 뿐만 아니라 동적인 특징을 고려하는 경우로 시간 지연 신경 회로망(TDNN : Time-Delay Neural Network)[19]과 재귀 신경 회로망 (Recurrent Neural Network)[17][18][20]이 대표적이다. 현재 주목을 받고 있는 재귀 신경 회로망은 OCON (One Class One Network)구조[22]로 이루어져 있으며 유

사한 음향학적인 특징을 가지는 음성 패턴의 클래스에 속하는 인식 단위에 대하여, 재귀 신경 회로망은 잘 정의된 학습 과정이 존재한다. 또한 재귀 신경 회로망은 인식 단위가 여러 개의 세그먼트로 이루어져 있으며 세그먼트 간의 경계가 모호한 경우에 적합하다. 따라서 인식 단위는 음절 혹은 단어 단위가 되며 인식 단위 내의 세그먼트의 경계는 출력 뉴런의 활성화 상태를 검사해 봄으로써 세그먼트간의 경계를 추출할 수 있다.

2.2 재귀 신경 회로망의 구조

재귀 신경 회로망의 구조는 그림 1과 같으며, n번째 뉴런의 활성 레벨 $y_n(t)$은 다음과 같이 주어진다.

$$y_n(t) = f_n\left[\sum_{l=0}^{N} w_{nl}\, y_l(t-1) + \sum_{m=0}^{M} w_{nm}\, u_m(t)\right] \qquad (2\text{-}1)$$

여기서, $u_m(t)$는 시간 t에서의 m번째 입력 단위이다. 그리고 w_{nl}은 l번째 뉴런에서 n번째 뉴런에 재귀적으로 연결된 weight이다. 또한 w_{nm}은 m번째 입력에서 n번째 뉴런에 연결된 가중치(Weight)이다. 그리고 $f_n(\bullet)$는 시그모이드 함수 이다. 입력에서 뉴런에 연결된 전향 가중치(Feed-forward weight)는 개개의 인식 단위 내에 있는 세그먼트의 정적 특징을 인식하기 위하여 학습하고, 반면에 뉴런 사이의 후향 가중치(feed-back weight)는 세그먼트의 동적 특징의 시간적인 변화를 특징짓기 위하여 사용된다. 만일 인식 단위 Γ가 K개의 세그먼트를 가지면, 이 인식 단위를 위한 재귀 신경 회로망의 모델은 K개의 출력 뉴런을 가지게 된다. 예를 들어 $u(t)$가 인식 단위 Γ를 모델링하기 위한 시간 t에서의 분석 프레임의 특징 벡터라고 하고 Γ의 k번째 세그먼트에 해당한다고 하면, k번째 출력 뉴런은 활성화되고 나머지 뉴런은 비활성화 된다. 즉,

$$y_k(t) = 1$$
$$y_k(t) = 0, \qquad \text{for all } 1 \le j \le k \text{ and } j \ne k \qquad (2\text{-}2)$$

이다.

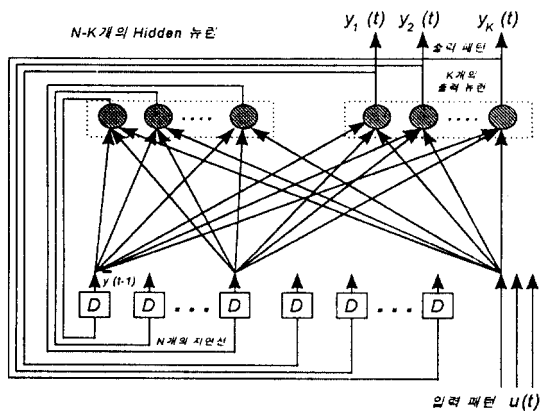

그림 1. 재귀 신경 회로망의 구조
Fig. 1 The structure of Recurrent Neural Network.

또한 $s(t)$가 시간 t에서의 활성화 상태를 나타낸다고 하면, 재귀 신경 회로망으로부터 다음과 같은 활성화 상태 열{$s(1)$, $s(2)$, ..., $s(T)$}을 얻을 수 있다.

$$s(t) = \arg \max_{i=1}^{K} \{y_i(t)\} \qquad (2\text{-}3)$$

또한 K개의 세그먼트의 시간적인 관계는 다음과 같다.

$$(1)\, s(t_1) \le s(t_2) \quad \text{if } t_1 \le t_2$$
$$(2)\, s(t+1) - s(t) = 0 \quad \text{or} \quad s(t+1) - s(t) = 1$$
$$(3)\, \text{For all } 1 \le k \le K$$
$$\text{there exists } t \text{ such that } s(t) = k \qquad (2\text{-}4)$$

그리고 k 번째 세그먼트의 시작을 $\tau(k)$라고 하면 다음과 같이 정의된다.

$$\tau(k) = \min\{t \mid s(t) = k\} \qquad (2\text{-}5)$$

2.3 재귀 신경 회로망의 학습 방법

재귀 신경 회로망의 학습은 재귀 신경 회로망의 weight를 조절하기 위하여 다음과 같이 정의된 시간 에러(Temporal Error) 함수에 역전파 학습 규칙[23]을 적용한다.

$$E = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{K}\sum_{k=1}^{K}[d_k(t) - y_k(t)]^2 \qquad (2\text{-}6)$$

여기서, $d_k(t)$는 원하는 출력이고, $y_k(t)$는 재귀 신경 회로망의 실제적인 출력을 나타낸다. 재귀 신경 회로망의 학습은 두 가지 방식으로 가능하다. 첫번째 방식은 실시간 처리를 위한 학습 방식이고, 두번째 방식은 기존의 역전파 학습 방식을 이용하는 것이다. 실시간 처리가 가능하도록 하는 방식의 경우 시간 0에서의 재귀 신경 회로망의 출력은 0이며, 이 때 출력은 연결된 가중치에 대하여 독립이라는 경계 조건하에서 학습을 수행한다.

$$\frac{\partial y_i(0)}{\partial w} = 0, \qquad i = 1, 2, \ldots, K \qquad (2\text{-}7)$$

따라서 현재 입력 프레임에 대하여 다음과 같은 가중치의 개선(Update)을 수행하면 된다.

$$w_{kl}(t+1) = w_{kl}(t) - \varepsilon \sum_{i=1}^{K} \mu_i(t)[y_i(t) - d_i(t)]\frac{\partial y_i(t)}{\partial w_{kl}} \qquad (2\text{-}8)$$

여기서 $\mu_i(t)$는 마스크 함수로 현재 뉴런에 연결된 가중치를 학습에 참여시 킬 것인가 제외시킬 것인가를 결정하는 함수이다. 그러므로 출력 노드의 경우 마스크 함수의 값은 1로 설정되며, 재귀 노드에 대해서는 마스크 함수가 0이 된다. 그리고 $d_i(t)$는 출력 노드의 예상 출력치(Target value)를 나타내고, $y_i(t)$는 현재 입력에 대한 신경 회로망의 실제적인 출력을 나타낸다. ε은 학습률 값(Learn-

ing rate)을 나타낸다. 또한 $\frac{\partial y_i(t)}{\partial w_{kl}}$는 네트워크의 다이나믹스(Dynamics)에 대한 가중치의 변화 효과를 나타내는 함수로 다음과 같이 정의된다.

$$\frac{\partial y_i(t)}{\partial w_{kl}} = f'(x_i(t)) \left[\sum_{j=1}^{K} w_{jk} \frac{\partial y_j(t)}{\partial w_{kl}} \right], \quad i=1, 2, ..., K \quad (2\text{-}9)$$

여기서 $f'(\bullet)$는 시그모이드 함수의 도함수를 나타내는, K는 출력 노드의 개수를 나타낸다. 그리고 M은 입력의 개수를 나타낸다. 두번째로 기존의 역전파 학습 방법을 이용하는 방식은 시간 $T+1$에서의 재귀 신경 회로망의 시간 에러는 이 때의 출력에 대하여 독립적이라는 경계 조건 하에서 학습이 이루어진다.

$$\frac{\partial E}{\partial y_i(T+1)} = 0, \quad i=1, 2, ..., K \quad (2\text{-}10)$$

따라서 현재 입력 프레임에 대하여 다음과 같은 가중치 개선을 수행하면 된다.

$$w_{kl}(t+1) = w_{kl}(t) - \varepsilon \frac{\partial E}{\partial w_{kl}} \quad (2\text{-}11)$$

여기서 $\frac{\partial E}{\partial w_{kl}}$는 시간 에러의 기울기를 나타내며 다음과 같이 정의될 수 있다.

$$\frac{\partial E}{\partial w_{kl}} = \sum_{t=1}^{T} \frac{\partial E}{\partial y_k(t)} f'(x_k(t-1)) z_l(t-1) \quad (2\text{-}12)$$

여기서 $z_l(t)$는 신경 회로망의 입력을 나타내며 다음과 같이 정의 된다.

$$z_l(t) = \begin{cases} y_l(t), & w_{kl}\text{이 재귀 부분의 가중치일 때} \\ u_l(t), & w_{kl}\text{이 외부 입력의 가중치일 때} \end{cases} \quad (2\text{-}13)$$

2.4 음성 인식에 적용한 재귀 신경 회로망

재귀 신경 회로망을 이용하여 음성 인식을 수행하기 위해서는 우선 인식 단위를 선정하고, 선정된 인식 단위에 맞게 인식 대상 어휘에 대한 모델링을 수행하여야 한다. 재귀 신경 회로망에서의 인식 단위는 분리하기 어려운 개체(Entity)를 인식 단위로 선정하기 때문에 주로 단어, 또는 음절을 인식 단위로 선정하게 된다. 인식 단위의 선정 이후 인식 대상 어휘에 대한 모델링이 수행되어야 하는데, 이 모델링 과정은 재귀 신경 회로망의 학습을 통하여 이루어진다. 학습을 수행할 때 모델링하고자 하는 인식 단위에 대한 학습 데이터는 길이 T인 특징 벡터의 시간적인 열로 구성되어 있다. 이 시간적인 열을 {u(1), u(2), ..., u(t), ..., u(T)}로 정의한다. 예를 들어 한국어 숫자음 음절 /칠/(/c^hil/)을 모델링하는 과정을 살펴보면, 음절 /칠/(/c^hil/)의 특징 벡터의 열을 재귀 신경 회로망에 시간적인 순서에 따라 입력하게 된다. 만일 임의의 t번째 특징 벡터 u(t)가 입력되면 재귀 신경 회로망에서는 음절 모델에 속해 있는 모든 세그먼트에 대한 확률값을 출력한다. 이 확률값은 {y_1(t), y_2(t), ..., y_k(t), ..., y_K(t)}로 정의된다. 이 확률값 중에서 가장 큰 값이 y_k(t)라고 하면 입력된 특징 벡터 u(t)는 음절 모델의 k번째 세그먼트에 속하게 된다. 즉 상태 s(t)=k가 된다. 입력에 대한 상태 s(t)의 열{s(1), s(2), ..., s(T)}은 식(2-4)의 관계를 만족하도록 반복적으로 학습을 수행하게 된다. 인식 과정을 살펴보면, 미지의 음성으로부터 특징 벡터를 추출하는 전처리 과정을 거쳐 길이 T인 특징 벡터의 시간적인 열{u(1), u(2), ..., u(t), ..., u(T)}이 재귀 신경 회로망에 하나씩 순서적으로 인가되면, 1번째 특징 벡터 u(t)에 대하여 재귀 신경 회로망은 {y_1(t), y_2(t), ..., y_k(t), ..., y_K(t)}를 출력하게 되고, 이 중에서 가장 큰 값이 y_k(t)이면 s(t)=k가 된다. 모든 특징 벡터가 순차적으로 입력된 이후에는 상태열{s(1), s(2), ..., s(T)}에 대하여 식(2-4)의 관계를 만족하는지 여부를 판별하게 된다. 이 상태열에 대한 시간적인 관계를 만족하는 음절들 중에서 가장 큰 누적 확률값을 가지는 모델을 인식된 결과로 간주하게 된다.

2.5 기존의 재귀 신경 회로망의 문제점

기존의 재귀 신경 회로망을 음성에 적용했을 때의 문제점은 재귀 신경 회로망의 구조적인 문제점과 학습 방법의 문제점으로 나누어 볼 수 있다. 첫번째로 구조적인 문제점은 구조상 입력과 출력층만으로는 분류하고자 하는 클래스의 경계 영역이 곡선(Convex) 형태에 국한된다는 것이다. 예를 들어 분류하고자 하는 클래스의 경계 영역이 3차원 공간상에 임의의 폐곡선 형태를 가진다면, 입력과 출력층만을 가진 재귀 신경 회로망으로는 분류하고자 하는 클래스의 정확한 경계 영역을 분류하지 못한다. 음성의 경우 분류하고자 하는 클래스는 경계 영역이 초평면(Hyper-plane) 상에서 임의의 모양을 가진 폐쇄영역을 형성하기 때문에 입력과 출력층만을 가진 재귀 신경 회로망으로는 정확한 분류 능력을 기대하기 어렵다. 두번째로 학습 방법의 문제점은 재귀 신경 회로망의 학습에서 수순한 재귀 노드에는 출력 예상치가 주어지지 않으므로 재귀 노드는 학습에서 제외된다는 것이다. 수순한 재귀 노드를 학습에서 제외시켰을 경우 학습 시간이 길어지고, 학습 과정 중에 원하는 시간 에러 값으로 수렴하지 못하고 진동을 하는 경우가 발생한다. 신경 회로망의 진동은 최적의 가중치를 찾지 못한다는 것을 의미한다.

Ⅲ. 제안된 연속 숫자음 인식 시스템

3.1 개 요

연속 숫자음 인식을 위한 전체 인식 시스템 블럭도는 그림 2와 같다.

위의 전체 음성 인식 시스템은 음절 인식 부분과 음절 모델의 확률값을 가지고 최적의 음절열을 만들어 내는

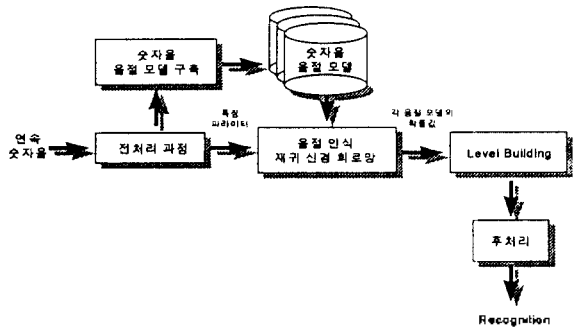

그림 2. 전체 음성 인식 시스템의 블럭도
Fig. 2 The Block Diagram of Speech Recognition System.

과정으로 나누어진다. 음절 인식 부분은 한국어 숫자음의 음절을 인식하는 부분으로 각 숫자음 음절 모델의 만드는 학습 과정과 미지의 숫자음 음절에 대한 세그먼트의 유사도를 출력하는 부분으로 구성되어 있다. 그리고 최적의 음절열을 만들어 내는 과정은 레벨 빌딩과 후처리 과정을 통해 이루어진다. 미지의 연속 숫자음이 본 시스템에 인가되었을 때 음절 인식 부분은 입력된 음성의 특징 벡터열에 대한 각 모델의 세그먼트의 유사도를 출력하고, 이 유사도를 레벨 빌딩 과정에서 거리값으로 변환하여 원하는 레벨 수에 대한 최적의 음절 열을 결정한다. 결정된 음절열에 후처리를 가하여 최적의 음절열을 출력하게 된다. 레벨 빌딩 알고리듬의 수행은 재귀 신경 회로망의 출력이 하나의 모델에 대하여 모델 내에 있는 세그먼트에서 각각 확률값을 출력하므로 각 모델에 대하여 적절한 상태 전이를 하고 있는지 여부를 평가하고 올바른 상태 전이를 하는 모델에 대하여 레벨 빌딩 알고리듬[24]을 수행한다.

3.2 제안된 재귀 신경 회로망

3.2.1 제안된 재귀 신경 회로망의 구조

기존의 재귀 신경 회로망은 입력과 출력의 두 층으로 구성되어 있기 때문에 분류하고자 하는 클래스의 경계 영역이 임의의 폐쇄영역을 형성하는 경우에 대하여 부적절하다. 따라서 재귀 신경 회로망의 구조적인 확장은 필수적이다. 제안된 재귀 신경 회로망의 구조는 그림 3과 같다. 그림 3(a)는 재귀 신경 회로망의 출력 부분만을 확장한 것이고, 그림 3(b)는 신경 회로망의 재귀 부분과 출력 부분을 모두 확장한 것이다. 그림 3(a)는 재귀 부분에 대해서는 확장하지않고 출력 노드 부분만을 확장함으로써 좀더 정확한 출력 상태열을 얻을 수 있으며, 이로 인하여 보다 나은 인식률을 얻을 수 있다.

그림 3(b)는 재귀 부분과 출력 부분을 모두 확장함으로써 좀더 정확한 상태열 및 세그먼트의 상태전이를 얻을 수 있다. 따라서 정확한 세그먼트의 경계 검출 및 상태 전이를 얻을 수 있으므로 각 모델에 대한 신경 회로망의 분류 능력을 높일 수 있다.

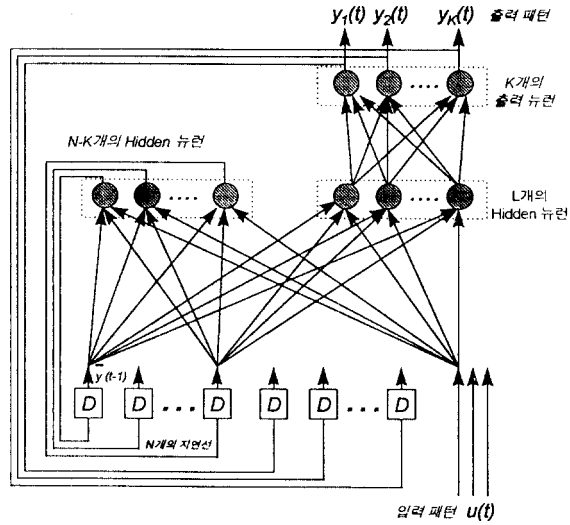

그림 3. 제안된 재귀 신경 회로망의 구조 (a) 출력 부분을 확장한 경우
Fig. 3 The structure of Expanded Recurrent Neural Network (a) Expanded Recurrent Neural Network in output parts.

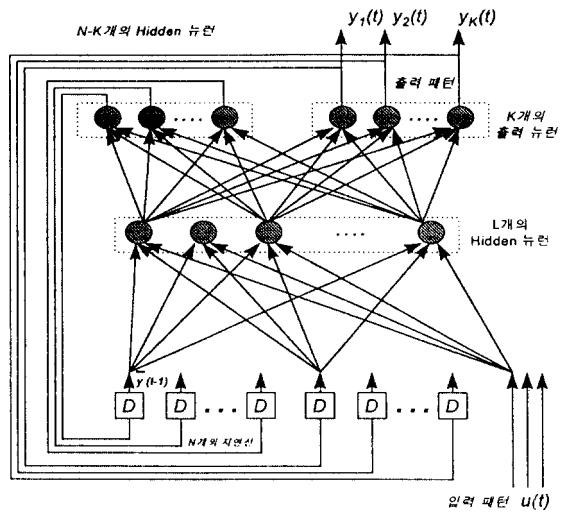

그림 3. 제안된 재귀 신경 회로망의 구조 (b) 재귀, 출력 부분을 확장한 경우
Fig. 3 The structure of Expanded Recurrent Neural Network (b) Expanded Recurrent Neural Network in output and recurrent parts.

3.2.2 제안된 재귀 신경 회로망의 학습 방법

제안된 재귀 신경 회로망의 학습 속도 향상을 위하여 자동 세그멘테이션(Self-Segmentation) 과정과 재귀 노드의 출력 예상치를 인가하는 두가지 방법을 제시한다.

3.2.2.1 자동 세그멘테이션 과정

기존의 재귀 신경 회로망을 학습시키는 두가지 방식이 모두 사전에 학습 단위의 세그먼트 정보를 알고 있어야

만 원하는 출력 상태열을 얻을 수 있다. 그러나 실제의 경우에 이것은 매우 비합리적이기 때문에 다음과 같이 반복적으로 자동 세그멘테이션을 수행하는 과정이 포함되어야만 한다. 만일 인식 단위 Γ를 학습시키기 위한 학습 데이터 그룹을 $\Omega=\{U^{(1)}, U^{(2)}, U^{(P)}\}$라고 가정하고, $U^{(P)}$에서 추출한 특징의 열을 $\{\vec{u}^{(P)}(t)\}$라고 하면, $\{\vec{u}^{(P)}(t)\}$를 재귀 신경 회로망에 인가하면 재귀 신경 회로망은 활성화 상태와 비활성화 상태의 두가지 상반된 반응을 보일 것이다. 이 반응열에 대하여 수식 (2-5)를 적용하여 유효한 상태열$\{s(t)\}$을 얻게 되면 원하는 출력 상태열을 재귀 신경 회로망 자체에서 다음과 같이 얻을 수 있다.

$$d_k(t)=\begin{cases}1.0 & , \text{ if } s(t)=k\\ 0.0 & , \text{ if } s(t)\neq k\end{cases} \tag{3-1}$$

하지만 재귀 신경 회로망이 유효하지 않은 상태열을 발생하면, 미리 가정된 상태열을 인가하여야 한다. 미리 가정된 상태열은 다음과 같이 정의된다.

$$d_k(t)=\begin{cases}1.0 & , \text{ if } k=\lfloor\{t/(\max frame/K)\}\rfloor\\ 0.0 & , \text{ if } k\neq\lfloor\{t/(\max frame/K)\}\rfloor\end{cases} \tag{3-2}$$

유효하지 않은 상태열을 발생하는 경우는 학습하고자 하는 모델에 대하여 최초의 학습 데이터가 인가되는 경우이다. 최초의 학습 데이터가 인가된 이후에는 추가의 학습 데이터가 인가되더라도 재귀 신경 회로망은 새로운 데이터에 맞는 상태열을 출력하게 된다. 그림 4는 재귀 신경 회로망의 학습 과정을 보이고 있다.

3.2.2.2 재귀 노드의 출력 예상치 인가

기존의 재귀 신경 회로망은 재귀 노드에 학습 출력치가 주어지지 않음으로써 재귀 노드는 재귀 신경 회로망의 학습 과정에서 제외된다. 재귀 신경 회로망의 학습에서 재귀 노드가 제외되면 신경 회로망의 학습 시간이 길어지고, 재귀 신경 회로망이 국부 최소값(Local Minimum)에 빠지는 경우 시간 에러가 임계치 이하로 감소하지 않는다. 이 경우에는 사용자가 국부 최소값에서 탈출할 수 있는 조건을 주어야만 한다. 하지만 실제의 경우 이것은 비합리적이다. 따라서 재귀 노드에 출력 예상치를 인가함으로써 학습 과정 중 국부 최소값에서 발생하는 문제를 해결하여 학습의 효율성을 높이고, 학습 속도 향상을 도모할 수 있다. 재귀 노드에 인가되는 출력 예상치로는 현재 입력 프레임이 속해있는 세그먼트의 평균치를 인가할 수 있고, 현재 입력 프레임이 시간 t의 특징 벡터라고 하면 시간 t+1에 해당하는 특징 벡터를 인가할 수 있다. 현재 프레임의 평균치를 인가하는 것은 재귀 신경 회로망이 좀더 정확한 세그먼트의 경계를 출력하도록 의도된 것이며, 시간 t+1의 특징 벡터를 인가하는 것은 재귀 신경 회로망으로 하여금 입력 프레임의 변화 특성을 학습하도록 의도된 것이다.

3.3 연음 현상을 고려한 연속 숫자음 모델

연음 현상을 고려한 숫자음 모델은 표 1에 나타나 있다[25][26][27][28]. 표 1에서 기본 모델에 추가된 숫자음 모델을 살펴보면 두가지의 형태로 나누어 볼 수 있다. 첫번째 경우는 선행 숫자음의 종성이 음가가 있고, 반면에 후행 숫자음의 초성이 음가가 없는 경우로, 선행 숫자음의 종

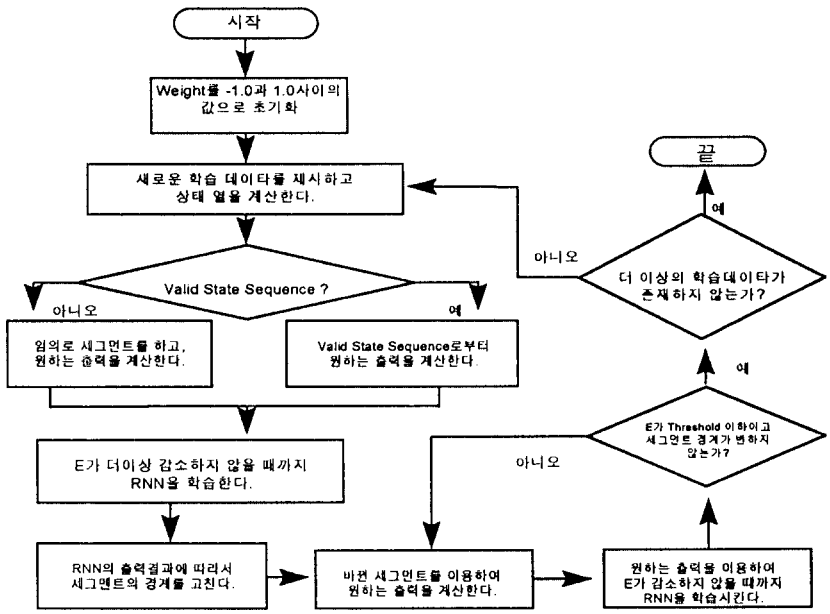

그림 4. 재귀 신경 회로망의 학습 과정
Fig. 4 The training procedure of Recurrent Neural Network.

성이 후행 숫자음의 초성 역할을 할 때 발생할 수 있는 음절을 모델링한 것이다. 두번째 경우는 선행 숫자음의 종성이 음가가 있고, 후행 숫자음의 초성 역시 음가를 가지는 경우로, 후행 숫자음의 초성의 음가만이 바뀌거나 선행 숫자음의 초성과 후행 숫자음의 종성의 음가가 동시에 바뀌는 경우가 이에 해당한다.

표 1. 연음 현상을 고려한 숫자음 음절 모델
Table 1. The prolongation in the continuous digits.

Index	1	2	3	4	5	6	7	8	9	10
모델	일 /il/	이 /i:/	삼 /sam/	사 /sa/	오 /o:/	육 /yuk/	칠 /c^{h}il/	팔 /phal/	구 /ku/	공 /koŋ/
Index	11	12	13	14	15	16	17	18	19	20
모델	ㄱ-일 /g-il/	ㄹ-일 /r-il/	ㅁ-일 /m-il/	ㄱ-이 /g-i:/	ㄹ-이 /r-i:/	ㅁ-이 /m-i:/	쌈 /s'am/	싸 /s'a/	ㄱ-오 /g-o:/	ㄹ-오 /r-o:/
Index	21	22	23	24	25	26	27	28		
모델	ㅁ-오 /m-o:/	ㄴ-육 /n-yuk/	ㄴ-용 /n-yuŋ/	ㄱ-구 /k-ku/	예 /e/	ㄱ-에 /g-e/	ㄹ-에 /r-e/	ㅁ-에 /m-e/		

3.4 연속 숫자음 인식을 위한 레벨 빌딩 알고리듬

레벨 빌딩 알고리듬은 연결어 및 연속어에 사용되는 대표적인 알고리듬으로 단독어 표준 패턴들을 미지의 연결어 및 연속어와 비교하여 일치하는 최적 단독어 열을 결정한다[29][30]. One-Stage DP에서는 단독어 표준 패턴을 템프리트(Template)로 사용하고 최소 거리값의 정합을 계산하여 최소의 누적 거리값을 가지는 템플리트의 열을 인식된 결과로 취한다. 이와 유사하게 본 연속 숫자음 인식 시스템에서는 각 숫자음 음절 모델을 템플리트로 사용하고, 각 숫자음 음절에서 출력되는 세그먼트에 속할 확률(Probability) 등을 거리값(Distance)으로 변환하여 변환된 거리값을 가지고 레벨 빌딩 알고리듬을 수행한다. 본 연속 숫자음 인식 시스템에 사용된 레벨 빌딩 알고리듬의 흐름은 다음과 같다.

(1) $P^v_1(t)$, $1 \le t \le T$: 레벨 1에서 표준 패턴 v에 대한 시험 패턴의 프레임 t 까지의 누적 거리값 (3-3)

(2) $F^v_1(t)$, $1 \le t \le T$: 레벨 1에서 시작점을 나타내는 포인터 (3-4)

각 레벨 l에서 최적 모델을 구하기 위하여 모델 v에 대한 최소화는 다음과 같은 알고리듬을 수행한다.

(3) $$P^B_l(t) = \min_{1 \le v \le V} P^v_l(t), \quad 1 \le t \le T \tag{3-5}$$

각 모델 중에서 최소의 거리값을 가지는 모델의 누적 거리값

$$W^B_l(t) = \arg \min_{1 \le v \le V} P^v_l(t), \quad 1 \le t \le T \tag{3-6}$$

최소의 누적 거리값을 출력하는 모델의 인덱스

$$F^B_l(t) = F^{W^B_l(t)}_l(t), \quad 1 \le t \le T \tag{3-7}$$

최적 음절 모델의 포인터

각각의 새로운 레벨은 이전의 레벨에 끝점 영역에서 초기 프레임을 선정하여 시작되고, 음절 모델과의 패턴 매칭에 의해 레벨 빌딩이 이루어진다. 이러한 레벨 빌딩 알고리듬은 원하는 레벨 수에 이를 때까지 반복적으로 수행되며, 각 레벨에서는 최소의 누적 거리값을 가지는 모델의 인덱스에 대한 포인터 $F^B_l(t)$가 얻어지고, 전체적인 최적열은 가능한 모든 레벨에 대해 최소 누적 거리값 $P^B_l(t)$로 부터 얻을 수 있다.

$$P^* = \min_{1 \le l \le L_{max}} [P^B_l(T)] \tag{3-8}$$

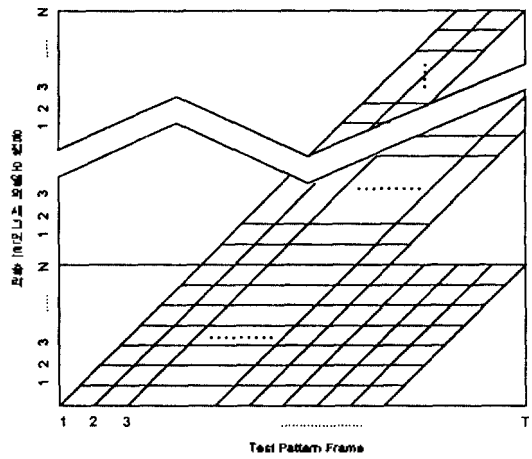

그림 5. 레벨 빌딩 과정
Fig. 5 Illustration on the Level Building Procedure.

3.5 후처리

연음 현상을 고려한 음절 모델 중 인덱스 11번에서 24번과 27번, 28번은 아래의 표 2와 같은 후처리 과정을 수행해야 한다.

표 2. 후처리
Table 2. The post-processing.

모델	선행 음절	선행음절 보정	후행음절 보정	모델	선행 음절	선행음절 보정	후행음절 보정
/g-il/ /g-i:/ /g-o:/ /k-ku/ /g-e/	/yuk/	/yuk/	/il/ /i:/ /o:/ /ku/ /e/	/r-il/ /r-i:/ /r-o:/ /r-e/	/il/, /i:/ /c^{h}il/ /p^{h}al/	/il/ /c^{h}il/ /p^{h}al/	/il/ /i:/ /o:/ /e/

/m-il/ /m-i:/ /m-o:/ /m-e/	/sa/	/sam/	/il/ /i:/ /o:/ /e/	/s'am/ /s'a/ /n-yuk/ /n-yuŋ/			/sam/ /sa/ /yuk/ /yuk/

Ⅳ. 실험 및 고찰

4.1 실험 환경

본 논문의 실험 환경은 다음과 같다. 음성 신호는 마이크에서 입력을 받아 TMS320C30보드에서 끝점을 검출한 이후 70Hz에서 4500Hz의 밴드 통과 필터를 거쳐 16bit A/D변환(10 kHz Sampling)을 수행하였다. 입력된 음성은 IBM/PC에서 LPC Cepstrum 계수를 구하여 본 인식 시스템에 인가되었다.

4.1.1 음성 데이터 베이스

음성 데이터 베이스의 구성은 다음 표 3와 같다. 단독 숫자음의 경우는 화자 종속에 대하여 5회씩 발음한 음성을 학습에 참여 시켰고, 나머지 15회의 발음은 인식 실험에 사용하였다. 화자 독립의 경우는 1인의 화자 음성을 학습에 사용하였고, 나머지 4인의 화자가 5회씩 발음한 10개의 단독 숫자음을 인식 실험에 사용하였다.

표 3. 음성 데이터 베이스의 구성
Table 3. The Configration of Speech Recognition DataBase.

구 분	화 자	화자수	대 상	발성 회수
단독 숫자음	화자 종속	1	10개의 숫자음	20
	화자 독립	4	10개의 숫자음	5
연속 숫자음	화자 종속	1	21개의 표준패턴	5
	화자 독립	4	21개의 표준패턴	5

연속 숫자음 인식의 경우는 연속 숫자음 사이에서 발생 가능한 연음 현상을 고려하여 만들어진 21개의 표준 패턴에 대하여 5회씩 발음한 음성을 인식에 사용하였다. 21 개의 연속 숫자음의 표준 패턴은 다음의 표 4와 같다.

표 4. 21 개의 연속 숫자음 표준 패턴
Table 4. 21 Reference Patterns of Continuous Digits.

인덱스	1	2	3	4	5	6	7
패턴	512-0257	630-1349	745-6780	826-9318	904-0371	910-2388	843-6416
인덱스	8	9	10	11	12	13	14
패턴	729-5522	607-7641	358-8736	153-0599	270-9483	396-0011	408-6281
인덱스	15	16	17	18	19	20	21
패턴	689-6542	209-1921	147-3324	986-5066	569-1775	795-9785	448-1234

4.1.2 특징 파라미터 추출

그림 6은 특징 파라미터 추출 과정을 보이고 있다. 특징 파라미터는 끝점이 검출된 음성을 70Hz-4500Hz의 대역 통과 필터로 필터링을 한 후 16bit A/D변환(10 kHz샘플링)한다. A/D 변환된 음성으로부터 10차의 LPC 계수를 구한 후 10차의 LPC Cepstrum 계수를 구하여 사용하였다. 한국어 단독 숫자음 11개를 사용한 화자 종속 인식 실험 결과, LPC Cepstrum 계수의 차수를 8차로 하였을 경우 89.3%, 10차로 하였을 경우 90.6%, 12차로 하였을 경우 90.6%의 인식률을 얻었다. 따라서 본 논문에서는 특징 파라미터로 10차의 LPC Cepstrum 계수를 사용하였다.

표 5. 특징 파라미터의 선정
Table 5. The Choose of Feature Parameter.

LPC Cepstrum 차수	8차	10차	12차
인식률	89.3%	90.6%	90.6%

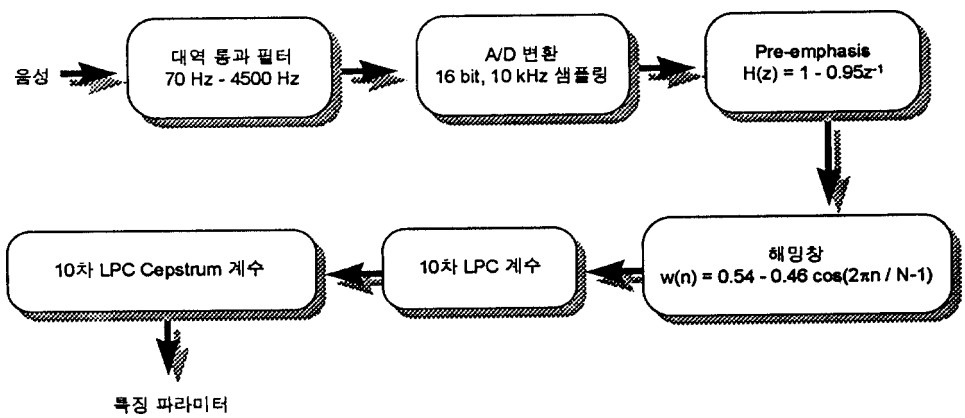

그림 6. 특징 파라미터의 추출
Fig. 6 The extraction of feature parameter.

4.2 인식 실험

인식률 향상을 위해 재귀 신경 회로망의 구조를 확장한 경우와 기존의 재귀 신경 회로망의 구조를 그대로 사용한 경우에 관한 실험과 숫자음 모델의 세그먼트수에 관한 실험 그리고 음절 모델의 수에 관한 실험, 다중의 음절 모델을 사용한 경우에 관한 실험으로 나누어 하였다.

4.2.2.1 재귀 신경 회로망의 확장에 관한 실험

다음의 표 6은 재귀 신경 회로망을 확장한 경우와 그렇지 않은 경우에 대한 실험 결과를 나타낸 것이다. 표 7은 단독 숫자음에 대한 화자 종속과 화자 독립에 대한 실험 결과이고 표 8은 연속 숫자음에 대한 실험 결과이다. 화자 독립의 경우는 4인의 화자가 각 숫자음을 5회씩 발음한 음성을 사용하였다.

표 6. 재귀 신경 회로망을 확장한 경우와 그렇지 않은 경우의 인식률

Table 6. The recognition rate of Expanded Recurrent Neural Network and Recurrent Neural Network.

단독 숫자음		/일/	/이/	/삼/	/사/	/오/	/육/	/칠/	/팔/	/구/	/공/	Total
2 Layer		80.0	86.6	100	86.6	86.6	100	86.6	100	86.6	93.3	90.6
3 Layer	(A)	93.3	86.6	100	86.6	86.6	100	86.6	100	86.6	93.3	92.0
	(B)	100	93.3	100	100	100	100	93.3	100	86.6	100	97.3

(A): 출력 부분만을 확장한 경우
(B): 재귀 부분과 출력 부분을 모두 확장한 경우

표 7. 단독 숫자음에 대한 화자 종속 및 화자 독립의 인식률

Table 7. The recognition rate of speaker dependent and speaker independent for the isolated digits.

	화자 종속	화자 독립
인식률	97.3%	80.5%

표 8. 연속 숫자음에 대한 화자 종속 및 화자 독립의 인식률

Table 8. The recognition rate of speaker dependent and speaker dependent for the continuous digits.

	화자 종속	화자 독립
인식률	88.2%	81.3%

4.2.2.2 연속 숫자음 모델 내의 세그먼트 수에 관한 실험

연속 숫자음 모델 내의 세그먼트에 관한 실험은 음절 모델이 초성, 중성, 종성으로 이루어져있기 때문에, 각각 초성, 중성, 종성을 하나의 세그먼트로 간주하여 각 음절을 3개의 세그먼트로 구성한 경우, 초성, 중성, 종성을 각각 2개의 세그먼트로 간주하여 각 음절을 6개의 세그먼트로 구성한 경우 그리고 초성, 중성, 종성을 각각 3개의 세그먼트로 간주하여 9개의 세그먼트로 구성한 경우에 대하여 수행하였다. 실험 결과, 각 음절 모델을 6개의 세그먼트로 구성하였을 경우가 가장 좋은 인식률을 나타내었다.

4.2.2.3 음절 모델 수에 관한 실험

음절 모델 수에 관한 실험은 11개의 음절 모델을 사용한 경우와 17개의 음절 모델을 사용한 경우, 연음 현상을 고려하여 28개의 음절 모델을 사용한 경우에 대하여 인식 실험을 수행하였다. 실험 결과, 17개의 음절 모델을 사용한 경우가 가장 높은 인식률을 나타내었다. 17개의 음절 모델은 11개의 음절 모델을 사용하여 실험한 결과를 분석하여 오인식이 많이 발생하는 음절에 대해서만 연음 현상을 고려한 것이다. 다음의 표 9는 음절 모델 수의 변화에 따른 화자 종속 연속 숫자음 인식에 관한 실험 결과이다.

표 9. 음절 모델의 수 변화에 관한 연속 숫자음 인식 실험 결과

Table 9. The recognition rate of Various Syllable Models.

음절 모델 수	11개의 모델	17개의 모델	28개의 모델
인식률	64.9%	79.2%	61.3%

그리고 표 10은 17개의 음절 모델을 표시하고 있다.

표 10. 17개의 음절 모델의 구성

Table 10. The configuration of 17 Syllable Models.

인덱스	1	2	3	4	5	6	7	8	9
모델	/il/	/i:/	/sam/	/sa/	/o:/	/yuk/	/c^{h}il/	/p^{h}al/	/ku/
인덱스	10	11	12	13	14	15	16	17	
모델	/koŋ/	/s'a/	/r-i:/	/m-i:/	/m-o:/	/g-o:/	/e/	/r-e/	

4.2.2.4 다중 음절 모델에 관한 실험

다중 음절 모델을 관한 실험은 11개의 음절 모델을 사용하였을 경우 오인식이 자주 발생하는 음절에 추가의 음절 모델을 첨가한 경우에 대한 실험 결과이다.

실험 결과, 다중 음절 모델을 사용하였을 경우가 17개의 음절 모델을 사용하였을 경우보다 높은 인식률을 나타내었다.

표 11. 다중 음절 모델을 사용한 경우의 인식률

Table 11. The recognition rate of Multiple Syllable Models.

모델 종류	17개의 모델	다중 모델
인식률	79.2%	86.3%

다음의 표 12는 다중 음절 모델의 구성을 나타내고 있다.

표 12. 다중 음절 모델의 구성

Table 12. The configuration of Multiple Syllable Models.

모델	수	모델	수	모델	수
/il/	1	/o:/	2	/ku/	1
/i:/	1	/yuk/	1	/koŋ/	1
/sam/	1	/c^{h}il/	2	/e/	2
/sa/	1	/p^{h}al/	3		

4.3 고 찰

재귀 신경 회로망을 학습시킬 때 자동 세그멘테이션 과정을 포함시키고 재귀 노드에 출력 예상치를 인가함으로써 일정한 반복 회수 이상에서 시간 에러가 위의 방법을 수행하지 않았을 경우 보다 낮다는 것을 보았는데, 이는 자동 세그멘테이션 과정과 재귀 노드에 출력 예상치를 인가 하는 것이 학습 속도를 향상 시킨다는 것을 입증하는 것이다. 재귀 노드에 출력 예상치를 인가할 때 입력 프레임이 속해있는 세그먼트의 평균치를 인가하는 것이 다음 프레임을 인가하는 것보다 훨씬 효율적이라는 사실도 알 수 있었다. 단독 숫자음의 인식 실험에서 화자 종속에 대하여 입력과 출력의 두 층만을 가지는 경우 90.6%, 출력 부분을 확장한 경우 97.3%의 인식률을 얻을 수 있었다. 실험 결과에서 보는 바와 같이 출력 부분과 재귀 부분을 모두 확장 시켰을 때 가장 좋은 결과를 얻을 수 있었다. 음절 모델의 세그먼트 수에 관한 실험에서 재귀 신경 회로망은 자동 세그멘테이션 과정을 통하여 세그먼트 수를 자동으로 줄이게 되며, 결국 세그먼트 수는 각 음절 모델에 대하여 최대 3개, 최소 1개가 된다. 하지만 이렇게 만들어진 음절 모델을 연속 숫자음 인식에 사용하면 인식 되어야 할 음절 모델의 세그먼트의 출력에서보다 다른 음절 모델의 세그먼트에서 더욱더 큰 유사도를 출력하는 경우가 빈번하여 현저한 인식률의 저하를 가져왔다. 특별히 hidden 노드에서의 에러 값에 기인한 인식률의 저하를 초래 하였다. 이러한 실험 결과를 토대로 각 음절의 모델의 초성, 중성, 종성을 2개내지 3개의 세그먼트로 각각 나누어 음절 모델을 구성하고 인식 실험을 한 결과 6개의 세그먼트를 가지는 경우가 가장 좋은 인식률을 나타내었다. 연속 숫자음의 인식의 경우는 11개의 숫자음 모델과 연음 현상을 모두 고려하여 만든 28개의 숫자음 모델, 그리고 오인식이 자주 발생하는 부분에 대해서만 연음 현상을 고려하여 만든 17개의 모델을 가지고 인식 실험을 한 결과, 오인식이 발생한 부분에 대하여 연음 현상을 고려하여 만든 모델이 가장 좋은 인식률을 보였다. 연음 현상을 모두 고려한 28개의 음절 모델과 오인식이 빈번히 발생하는 음절에 대해서만 연음 현상을 고려한 모델의 인식 결과는 모두 후처리를 수행하여 인식을 한 것이다. 그리고 오인식이 발생하는 부분에 대하여 연음 현상을 고려한 17개의 모델과 오인식이 빈번히 일어나는 음절에 대하여 중복된 모델을 첨가한 16개의 모델의 인식 실험 결과는 다중의 모델을 사용한 경우가 더욱더 높은 인식률을 보였다.

V. 결 론

본 논문에서는 재귀 신경 회로망의 학습 속도 향상을 위하여 자동 세그멘테이션 과정과 재귀 노드에 출력 예상치를 인가하는 방법을 제안하였다. 그리고 인식률 향상을 위하여 재귀 신경 회로망의 구조적인 확장을 제안하였으며, 연속 숫자음의 인식률을 높이기 위하여 여러 가지 모델을 제안하였다. 본 논문에서 제안한 학습 속도 향상 방법을 한국어 숫자음에 적용한 결과 학습 속도가 향상됨을 알 수 있었다. 또한 재귀 신경 회로망의 구조적인 확장을 수행하여 인식 실험을 한 결과, 구조적인 확장을 하지 않은 경우 보다 확장한 경우가 높은 인식률을 나타내었다. 이로서 재귀 신경 회로망의 구조적인 확장의 필요성을 입증할 수 있었다. 연속 숫자음의 인식률 향상을 위하여 여러 가지 모델을 제시하였으며, 실험 결과에서는 기본 음절 모델에 오인식이 빈번히 발생하는 음절의 모델을 중복하여 사용한 경우가 가장 좋은 인식 결과를 나타내었다. 본 논문에서 제안한 재귀 신경 회로망과 학습 속도 향상을 위한 방법은 음성 다이얼링 시스템과 같은 응용 시스템에 적용 가능하며, HCI (Human-Computer Interface) 기술 발전에 기여할 것으로 사료된다. 그리고 향상된 인식률을 얻기 위해서는 기존의 알고리듬에 대한 추가의 연구가 필요할 것으로 사료된다.

참 고 문 헌

1. Hiroaki Sakeo, Seibi, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. On Acoustics, Speech and Signal Processing, Vol. 26, No. 1, Feb, 1978.
2. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition" IEEE Proceeding, 1989.
3. R. P. Lipmann, "An Introduction to Computing with Neural Nets" IEEE ASSP Magazine, April, 1987.
4. J. W. Forgie and C. D. Forgie, "Results Abtained from a Vowel Recognition Computer Program" Journal of the Acoustical Society of America, Vol. 31, 1959.
5. B. Gold, "Word Recognition Computer Program" Technical Report 452, Research Laboratory of Electronics, M. I. T, Cambridge, MA. 1966.
6. M. R. Sambur and L. R. Rabiner, "A Speaker Independent Digit Recognition System" The Bell System Technical Journal, Vol. 54, No. 1, Jan. 1975.
7. D. R. Reddy, "Speech Recognition by Machine: A Review" Proceedings of the IEEE, Vol. 64, No. 6, April, 1976.
8. A. J. Gray, J. D. Marker, "Distance Measure for Speech Processing" IEEE Trans. On Acoustics Speech and Signal Processing, Vol. 24, No. 5, Oct. 1976.
9. B. Lowerre, "The Harpy Speech Understanding System" Trends in Speech Recognition, Prentice Hall, 1980.
10. Kai-Fu Lee, "Automatic Speech Recognition" Kluwer Academic Publishers, 1989.
11. Kai-Fu Lee, R. Reddy, "An Overview of the SPHINX Speech Recognition System" IEEE Trans. On Acoustics Speech and Signal Processing, Jan. 1990.
12. Kenji Kita, Wayne H. Ward, "Incorporation LR Parsing into SPHINX" ICASSP-91, 1991.

13. W. S. McCulloch, W. Pitts, "A Logical Calculus of The Ideas Immanent in Nervous Activity." Bulletin of Mathematical Biophysics, Vol. 5, 1943.
14. Donald O. Hebb, "The Organization of Behavior" New York, pp. 60-78.
15. F. Rosenblatt, The Perceptron: "A Probabilitic Model for Information Storage and Organization in The Brain" Psychological Review, Vol. 65. 1958.
16. M. Minsky, S. Papert, "Perceptron" Cambrige, MA:MIT Press, 1969.
17. M. D. Hanes, S. C. Ahalt and A. K. Krishnamurthy, "Acoustic to Phonetic Mapping using Recurrent Neural Networks" IEEE Trans. On Neural Networks, Vol. 5, No. 4, 1994.
18. S. J. Lee, K. C. Kim, H. Y. Yoon and J. W. Cho, "Application of Fully Recurrent Neural Networks for Speech Recognition" ICASSP 91, Vol. 1, 1991.
19. Alex Waibel, Hiderfumi Sawai and Kiyohiro Shikane, "Consonant Recognition by Modular Construction of Large Phoneme Time Delay Neural Networks" ICASSP 89, 1989.
20. Tan Lee, P. C. Ching, L. W. Chan, "Recurrent Neural Networks For Speech Modeling and Speech Recognition" EUROSPEECH 95, 1995.
21. C. H. Chen, "Fuzzy Logic and Neural Network Handbook", McGraw-Hill Inc., 1992.
22. S. Y. Kung, "Digital Neural Networks", Prentice Hall International Inc., 1993.
23. Rumelhart, D. E., G. E. Hinton, and R. J. Williams, "Learning Representations by Error Propagation Learning", Neural Computation, 1991.
24. C. S. Myers and L. R. Rabiner, "Connected Digit Recognition using a Level Building DTW Recognition", IEEE Trans. ASSP, Vol. 29, 1981.
25. 허 웅, "국어 음운학", 정음사, 1981.
26. 이 강성, "연결어 인식을 위한 음소 분류 신경 회로망과 LR 구문 분석법에 관한 연구", 광운대학교 전자 계산기 공학과, 1992.
27. 김 석득, 김 차균, 이 기백, "국어 음운론", 한국방송통신대학 출판부, 1981.
28. 남 기심, 고 영근, "표준 국어문법론", 탑출판사, 1994.
29. D. F. Specht, "Probabilitic Neural Networks", Neural Networks, Vol. 3, 1990.
30. L. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall International Inc., 1993.

▲이 성 권(Seong Kwon Lee)
음향학회지 15권 2E호 참조
현재:광운대학교 컴퓨터공학과 박사수료
※주관심분야:음성인식, 멀티미디어, 신호처리

▲김 순 협(Soon Hyob Kim)
음향학회지 15권 2E호 참조
현재:광운대학교 컴퓨터공학과 교수
※주관심분야:음성언어 및 합성, 다지털신호처리, 신경망