論文 96-21-7-1

시공간 신경회로망을 이용한 연속 숫자음 인식

學生會員 李 鍾 植* 正會員 鄭 在 皓*

Continuous Digits Recognition Using Spatio-Temporal Neural Network

Jong Sik Lee* *Student Member* Jae Ho Chung* *Regular Member*

요 약

본 논문에서는, 시공간 신경회로망을 이용하여 한국어 연속 숫자음의 인식을 시도하였다. 연속적으로 발음한 7자리 숫자음을 인식 목표로 하였으며, 7자리 연속음의 초기 인식률은 28%로 저조하였다. 본 논문에서는, 시공간 신경회로망에서 연속음의 인식률 향상을 위해 두 가지 방법을 제안하였다. 첫번째 제안점은 한국어 숫자음의 발음특성에 맞추어, 시작점은 에너지와 zero-crossing rate의 비교로써 검출하고, 끝점은 에너지 만으로 검출하였다. 이와 같은 숫자음의 발음 특성을 고려한 시작점, 끝점 검출방법을 사용하여, 7자리 연속음의 인식률을 61%로 향상시켰다. 연속음의 인식률 향상을 위한 두번째 제안점은, 같은 단어라 하더라도 연속음의 경우 발음이 앞뒤 단어들의 영향을 받아 변화하는 것을 고려하여, STNN의 갯수를 확장하는 것이다. 즉, 각 단어를 나타내는 STNN의 갯수를 1개에서 5개로 늘려서, 같은 단어가 여러 가지로 달리 발음되어도, 충분히 그 다른 발음의 특징을 별도로 구별된 STNN에서 인식할 수 있게 하였다. 이로써, 7자리 연속음의 인식률을 89%로 향상시켰다.

ABSTRACT

In this paper, a new approach for continuous digits recognition using the Spatio-Temporal Neural Network (STNN) is reported. The continuous seven digits are targeted to recognize, and our initial recognition rate was 28%. In this paper, to increase the recognition rate, two methods are proposed. In the first method, to compensate the STNN's own defect as well as to emphasize the Korean digits' phonic characteristics, the starting point of each digit is detected using the energy and zero-crossing rate, but the ending point is detected only using the energy value. In this case, the seven digits recognition rate increased to 61%. Furthermore, in the second method, con-

*인하대학교 전자공학과 디지탈 신호처리 연구실
Department of Electronic Engineering, Digital Signal Processing Laboratory, Inha University
論文番號:95243-0714
接受日字:1995年 7月 14日

sidering the fact that a same digit could be pronounced differently in continuously spoken environment, the number of STNNs used to represent each digit is increased from one to five. Consequently, the same digit but pronounced differently could be handled well in the new system. As a result of that, the continuously spoken seven digits recognition rate increased to 89%.

Ⅰ. 서 론

인간과 기계 사이의 보다 편리한 인터페이스의 필요성이 증가하면서, 인간과 기계의 의사 소통을 보다 자연스럽고 정확하게 하고자 하는 욕구가 늘고 있다. 통신, 신호처리 기술 등의 급속한 발달에 따라 인간의 가장 기본적인 통신 수단인 음성을 이용하는 것이 그 하나의 방법으로 주목받고 있으며, 음성을 이용한 사람과 기계의 인터페이스를 이루기 위한 핵심 기술이 음성 인식이다. 음성 인식이란, 인간의 음성을 인식알고리듬을 통해 단어나 문장으로 전환하는 것인데, 현재까지의 음성 인식 방법을 크게 두 가지로 나누면, 음성의 언어학적인 분법을 기반으로 이용하는 지식기반 접근방식과 입력된 음성 신호에서 특징값을 추출한 후 미리 준비된 기준 패턴과 비교하여 가장 유사한 것을 인식하는 패턴 매칭 방법으로 나눌 수 있다. 본 논분에서는 인식 방법으로서 패턴 매칭 방법의 하나인 신경회로망을 사용하였다.

신경회로망은 인간의 신경 조직을 모방하여, 신경 조직의 뉴런이 하는 기본 기능을 수행하는 요소들이 병렬로 상호 연결되어 있어서, 대량의 복잡한 데이타를 병렬처리할 수 있다. 그 뿐 아니라 학습능력이 있다는 사실에 근거를 두고, 신경망을 이용한 음성인식에 대한 연구가 활발히 진행되고 있다[1, 2, 3, 4, 5]. 그러나, 기존의 신경망들, 특히 일반화된 MLP(Multi-Layer Perceptron)나 CPN(Counter Propagation Network)은 정적 패턴의 인식에는 우수한 성능을 보이지만, 음성신호와 같이 시간에 따라 그 특성이 변하는 동적 패턴의 인식에는 취약점을 갖고 있는 것으로 알려져 있다. 이와 같은 음성신호의 특성, 즉 시간에 따라 순차적으로 변화하는 신호의 동적 특성을 효과적으로 인식할 수 있는 신경회로망으로서, 시공간 신경회로망 즉 STNN(Spatio-Temporal Neural Network)이 1980년대 중반 Kosko와 Klopf에 의하여 제안되었다[6, 7, 8, 9].

음성신호는 시간에 따라 특성이 변화하는 현상이외에, 같은 화자에 의하여 발음된 같은 음성신호라 하더라도 발음속도가 때에 따라 일정하지 않다. 또한 문맥의 전후 관계에 의하여 같은 음성이 조금씩 다르게 발음된다. STNN은 입력되어 들어오는 신호의 시퀀스 패턴들 상호간의 연관성을 고려하면서 인식을 하기 때문에, 위에 언급한 음성신호의 부분적인 시간 변화와 주파수 변화의 영향을 인식과정에서 감소시킬 수 있는 특징을 갖고 있다. 또한, STNN은 인식과정에서 화자의 발성 길이의 20% 정도의 증감은 인식결과에 크게 영향을 미치지 않는다는 큰 장점을 갖고 있다. 본 연구에서는 STNN을 사용하여 한국어 숫자음의 인식을 시도하였다.

본 논분은 한국어 연속 숫자음 인식에 대한 연구로서, 100개의 서로 다른 7자리 전화 번호를 연속적으로 발음한 데이타를 사용하여 인식 실험을 하였다. 연속적으로 발음된 7숫자음의 시작점과 끝점 검출을 위하여 에너지값과 zero-crossing rate을 사용하였다. 7자리 연속음의 초기 인식률은 28%로 저조하였다. 본 논분에서는, 연속음의 인식률 향상을 위하여 두 가지 제안을 하였다. 첫번째는, 한국어 숫자음의 발음 특성에 맞추어, 시작점과 끝점 검출을 시도하였다. 이를 위하여, 시작점은 에너지값과 zero-crossing rate의 비교로써 검출하고, 끝점은 에너지값 만으로 검출하였다. 또한, 본 논분에서 제안된 시작점 끝점 검출법은, 마지막 몇 개의 가중치 구간들과 입력 구간들의 유사성이, 처음과 중간의 가중치 구간들과 입력 구간들의 유사성보다 상대적으로 최종 출력값에 큰 영향을 미치는 STNN 자체의 미비점을 보완할 수 있었다. 한국어 숫자음의 발음특성에 맞추어진, 시작점과 끝점의 검출로 7자리 연속음의 인식률은 61%로 향상되었다. 그리고, 두번째는, 각 숫자음을 나타내는 STNN의 갯수를 증가시켰다. 즉, 같은 단어일지라도, 연속음의 경우 발음이 여러 가지로 변하는 경우를 생각하여 각 단어의 STNN의 갯수를 1개에서 5개로 늘리서, 같은

단어가 여러 가지로 달리 발음되어도, 충분히 그 다른 발음을 별도로 구별된 STNN에서 소화할 수 있게 하였다. 이와 같은 경우 7자리 연속숫자음의 인식률은 89%로 향상되었다.

2절에서는 STNN의 기본구조에 대하여 언급하였다. 3절에서는 본 논문에서의 제안점과 그 결과에 대하여 논하였다. 마지막으로 4절에서는 본 논문의 결론 및 향후 연구과제들에 대하여 언급하였다.

Ⅱ. 시공간 신경 회로망

시공간 신경 회로망은 여러 개의 층으로 구성되어 있으며, 각 층은 서로 다른 단어에 대한 가중치(weight)가 connection line들을 통하여 연결된 뉴런들로 구성되어 있다. 각 층의 구조는 동일하다. 각 층의 뉴런의 수는 입력 신호 전체의 길이를 선형적으로 나눈 구간의 갯수와 같다. 한 단어가 입력으로 들어오면, 시간적인 순서에 따라 첫번째 구간의 특징벡터 계수들이 층 전체의 뉴런들을 활성화시키고, 출력값을 낸다. 시간이 지남에 따라 두번째 구간의 입력신호가 들어오면, 두번째 구간의 특징벡터 계수들을 계산하고, 다시 층 전체의 뉴런들을 활성화하여 출력값을 계산한다. 이와 같은 과정을 마지막 입력신호의 구간까지 반복 수행하여, 최종 출력값을 구한다. 각 층의 동작 원리가 그림 1에 설명되어져 있으며, 각 층을 구성하고 있는 i번째 뉴런의 입력값을 수식적으로 나타내면 식 (1)과 같다.

$$I_i = \overline{Q_j} \cdot \overline{W_i} + d \sum_{k=1}^{i-1} x_k \qquad (1)$$

식 (1)에서, I_i는 i번째 뉴런의 입력값을, $\overline{Q_j}$는 j번째 입력구간의 특징벡터를, $\overline{W_i}$는 i번째 뉴런의 가중치 벡터를, x_k는 k번째 뉴런의 출력값을 나타낸다. 또한, 식 (1)에서 $d(<1)$는 i번째 뉴런에 전달되는 i번째 이전 뉴런들의 출력량의 크기를 조절하는 상수이다.

한편, i번째 뉴런의 출력값 x_i는 다음과 같은 미분방정식을 통하여 나타내어진다.

$$\dot{x}_i = A(-ax_i + b[I_i - \Gamma]^+) \qquad (2)$$

식 (2)에서, a와 b는 양의 상수값이다. 식 (2)에 포함되어 있는 함수 $[I_i - \Gamma]^+$는 다음과 같이 정의된다. 즉,

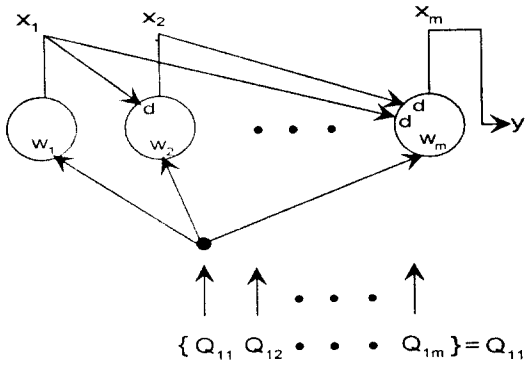

그림 1. STNN을 이용한 각 층의 구조

$$[I_i - \Gamma]^+ = \begin{cases} I_i - \Gamma & , \quad \text{if } I_i - \Gamma > 0 \\ 0 & , \quad \text{if } I_i - \Gamma \le 0 \end{cases} \qquad (3)$$

따라서, 식 (3)에서 Γ는 임계값(threshold)의 역할을 한다. 식 (2)에서 함수 $A(\cdot)$는 attack function이라 불리우며, 다음과 같이 정의된다.

$$A(u) = \begin{cases} u & , \quad \text{if } u > 0 \\ cu & , \quad \text{if } u \le 0 \end{cases} \qquad (4)$$

Attack function은 식 (2)에 소개한 i번째 뉴런의 출력값이, 균형상태(equilibrium state)에 이르기까지의 상승시간(rising time)과 균형상태 후의 하강시간(falling time)의 길이를 조정하는 효과를 갖는다. 특히, 파라메타 c $(0<c<1)$는 이전 뉴런들의 가중치 벡터들이, 대응하는 입력구간들의 입력 벡터들과 일치되었을 경우, 이러한 기억들이 현재 뉴런의 출력값에 영향을 미치게 하는 특성을 지닌다. 따라서, 입력신호의 일부에 잡음이 섞이거나, 발음이 다소 변하였더라도 이를 극복할 수 있게 된다.

그림 2에는 attack function의 시간에 따른 변화, 즉 neuron의 시간에 따른 출력값이 설명되어져 있다.

STNN의 인식 과정을 살펴보면, 시간이 완료되었을 때에, 마지막 뉴런의 최종 출력값이 입력신호와 입력신호를 적용한 STNN 사이의 닮은 정도를 나타낸다. 따라서, 입력된 숫자음을 인식하기 위하여서는,

입력신호의 특징벡터를 서로 다른 숫자음을 나타내는 STNN들, 즉 0부터 9까지 각각의 숫자음에 해당하는 STNN에 각각 입력신호로 적용하고, 10개의 최종 출력값을 얻은 후, 이들을 비교하여 최종 winner를 결정한다.

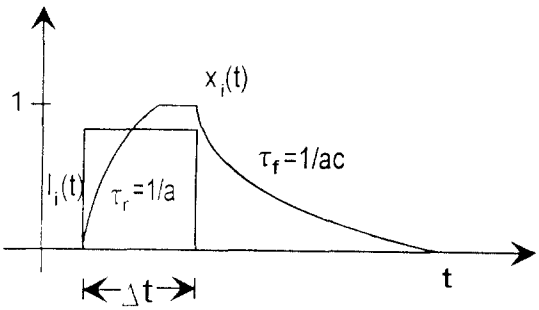

그림 2. 뉴런의 시간에 따른 출력값의 변화

시공간 신경 회로망의 기본 작동에 대한 내용은 참고문헌 [9,10]에 더욱 자세히 설명되어져 있다.

Ⅲ. 제안점과 실험 결과

3.1 실험 환경

실험은 화자 종속 시스템이며, 한 명의 남자 화자가 일주일 간격으로 3번, 100개의 서로 다른 7자리 전화 번호를 연속적으로 발음한 데이타를 사용하였다. 처음 두 주 사이에 얻은 200개의 전화 번호, 총 1400개의 숫자음을 가지고 신경망을 학습하였으며, 마지막 셋째 주의 100개의 전화 번호, 총 700개의 숫자음을 가지고 인식 실험을 하였다.

3.2 초기 연속음 인식률

본 연구에서 구성한 초기 음성 인식 시스템은 그림 3과 같다. 마이크를 통해 받아들여진 음성은 14bit 양자화 레벨을 갖는 A/D converter를 통하면서, 표본화 주파수 10KHz로 샘플링된다. 디지탈로 바뀐 음성신호에 Rabiner와 Sambur가 제안한 에너지값과 zero-crossing 알고리듬[11,12]을 적용하여 시작점과 끝점 검출을 한다. 시작점과 끝점이 검출된 신호는 시간 영역에서 전체의 길이를 10개의 선형적 프레임으로

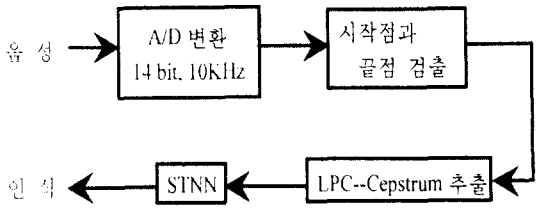

그림 3. 초기 음성 인식 시스템

나누어 분석한다. 각 프레임으로 부터 autocorrlation 방법을 이용한 Durbin 알고리듬을 통해서 16차 LPC 계수를 추출한다[11,13]. 계산된 LPC 계수들로부터 식 (5)와 식 (6)를 사용하여 16차 LPC-cepstrum 계수들을 계산한다. 즉, 본 논문에서는 입력 신호의 특징벡터로서 LPC-cepstrum을 사용하였다.

$$c(1) = -a(1), \tag{5}$$

$$c(n) = -a(n) - \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a(k)\, c(n-k), \quad 1 < n \le p. \tag{6}$$

식 (5)와 식 (6)에서 $a(n)$은 LPC 계수, $c(n)$은 LPC-cepstrum 계수, 그리고 p는 LPC 계수의 차수를 나타낸다. 마지막으로, 추출된 LPC-cepstrum 계수들을 STNN의 입력으로 사용하기 위해서 0과 1 사이의 값들로 정규화한다.

표 1은, 에너지값과 zero-crossing rate을 사용하여, 시작점과 끝점을 검출한 데이타로 실험한 결과를 보여 주고 있다. 각 숫자에 대해서는 82%의 인식률을 보였으며, 0과 9, 5와 9, 그리고, 1과 7사이에 많은 오

표 1. 7자리 연속음의 초기 인식률

	0	1	2	3	4	5	6	7	8	9	error 수	인식률
0	36									34	34	51.4%
1		64						6			6	91.4%
2			70									100%
3				69	1						1	98.5%
4					70							100%
5						39				31	31	55.7%
6							70					100%
7		54						16			54	22.8%
8									70			100%
9										70		100%
총 Error 수/총 인식률											126	82.0%

류를 보이고 있다. 그리고, 7자리 연속음이 모두 맞았을 때의 초기 인식률은 28%로 저조함을 보였다.

3.3 한국어 숫자음의 특성에 맞춘 끝점 검출(실험 Ⅰ)

실험 Ⅰ에서는, 한국어 단어의 발음 특성에 맞추어, 시작점과 끝점을 검출하여, 인식 실험을 하였다. 한국어 단어의 발음 특성을 살펴보면, 음의 앞부분에는 구개음, 마찰음, 파열음, 파찰음 등의 무성자음과 모음이 올 수 있고, 음의 끝부분에는 구개음을 포함한 휴지음(ㄱ, ㄷ, ㅂ)과 유성자음(ㄴ, ㅁ, ㅇ, ㄹ)으로 분류되는 7개의 종성 대표음과 모음이 올 수 있다. 특히, 한국어 숫자음(공, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구)은 표 2와 같이 분석된다.

표 2. 한국어 숫자음의 분석

음의 앞부분	음의 끝부분
구개음 : ㄱ	구 개 음 : ㄱ
마찰음 : ㅅ	유성자음 : ㅇ, ㄹ, ㅁ
파열음 : ㅍ	모 음 : ㅣ, ㅏ, ㅗ, ㅜ
파찰음 : ㅊ	
모 음 : ㅣ, ㅗ, ㅠ	

단어의 발음 특성을 살펴볼 때, 음의 앞부분은 에너지값의 비교와 zero-crossing rate의 비교로써 검출이 용이함을 알 수 있다. 반면에, 음의 끝부분은 마찰음, 파열음, 파찰음이 없으므로, 검출 방법에서 zero-crossing rate의 사용이 부적합하며, 에너지값의 비교가 검출에 용이함을 알 수 있다. 특히, 연속음들은 음들 사이가 가까이 붙어 있어, 다음에 오는 숫자음의 초성에 마찰음, 파열음, 파찰음 등이 오면, 음의 끝부분에서의 zero-crossing rate의 비교는 부정확한 끝점 검출의 원인이 된다. 즉, 연속음을 각 단어로 구분할 때는, 한국어 숫자음의 발음 특성에 맞추어, 시작점과 끝점을 검출하므로 인식률을 향상시킬 수 있게 된다. 따라서, 실험 Ⅰ에서는 시작점은 에너지값과 zero-crossing rate의 비교로써 검출하고, 끝점은 에너지값 만으로 검출하여, 인식 실험을 하였다.

또한, 한국어 단어의 발음 특성에 맞추어진 끝점 검출법은 동시에 STNN의 미비점을 보완하고 있다. STNN에서는, 시간적으로 앞에 있는 뉴런의 출력값은 바로 다음 뉴런의 출력값에 영향을 미치도록 구성되어 있으나, 뉴런의 출력값은 앞에 있는 뉴런의 출력값보다 그 뉴런에 해당되는 가중치 구간과 입력 구간의 유사성에 더 많은 영향을 받는다. 즉 식 (1)에서 $\overline{Q_j} \cdot \overline{W_i}$가 $d \sum_{k=1}^{i-1} x_k$ 보다 시간의 흐름에 따라 상대적으로 크게 된다. 증가된 i번째 뉴런의 입력값 I_i는 식 (2)를 통해 i번째 뉴런의 출력값 x_i를 크게 한다. 그러나, x_i는 d와 곱해져서 $d \sum_{k=1}^{i-1} x_k$의 일부분으로 반영되므로 I_{i+1}에 미치는 영향이 감소된다. 시간이 경과할수록 이 과정이 반복되어, 앞쪽에 있는 뉴런들의 출력값보다 뒤쪽에 있는 뉴런들에 해당되는 가중치 구간과 입력 구간의 유사성이 최종 출력값에 많은 영향을 미치게 된다. 결국, 마지막 몇 개의 가중치 구간들과 입력 구간들의 유사성은 처음과 중간의 가중치 구간들과 입력 구간들의 유사성 보다 상대적으로 최종 출력값에 큰 영향을 미치게 된다. 즉, STNN에서는 시작점 보다는 정확한 끝점의 검출이 인식률에 큰 영향을 미치게 된다. 따라서, STNN의 입력 패턴이 연속음인 본 실험에서, 초성에 마찰음, 파열음, 파찰음이 오는 경우에 끝부분에서의 zero-crossing rate의 비교를 하지 않으므로 더욱 정확한 끝점 검출을 할 수 있다. 이에 따라 끝점의 검출이 인식률에 많은 영향을 미치는 STNN의 미비점을 보완할 수 있다.

본 실험 Ⅰ에서 구성한 인식 시스템의 구조가 그림 4에 설명되어져 있다. 표 3은, 실험 Ⅰ의 결과를 보여주고 있다. 새롭게 끝점을 검출함으로, 0과 9, 5와 9, 그리고, 1과 7 사이에서 현저히 오류를 감소시켰으며, 각 단어에 대한 인식률을 93%로 증가시켰다. 또한, 7자리의 연속음이 모두 맞았을 때의 인식률은 61%로 향상되었다.

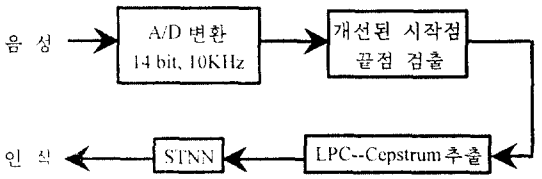

그림 4. 실험 Ⅰ에서 구성한 인식 시스템

표 3. 실험 Ⅰ의 인식률

	0	1	2	3	4	5	6	7	8	9	error 수	인식률
0	53					16				1	17	75.7%
1		70										100%
2			70									100%
3				67	3						3	95.7%
4					70							100%
5						70						100%
6							70					100%
7		4						66			4	94.2%
8									70			100%
9						25				45	25	64.2%
총 Error 수/총 인식률											49	93.0%

3.4 연속 숫자음의 특성에 맞춘 STNN의 확장(실험 Ⅱ)

실험 Ⅱ에서는, 실험 Ⅰ에서 제시한 방법으로 시작점과 끝점을 검출하고, 각 단어의 STNN의 수를 늘려서, 새로운 network을 구성한 후 인식 실험을 하였다. 실험 Ⅰ에서 사용된 STNN의 구성을 살펴보면, 각 단어에 대하여 1개의 STNN을 사용하였다(예:'0'의 STNN, '1'의 STNN, ...). 하지만, 같은 의미의 단어일지라도, 연속 숫자음의 경우, 발음이 여러 가지로 변하는 경우를 생각할 수 있다. 예를 들면, '6'의 발음은 '육', '융', '뉵', '륨' 등으로 각각 발음될 수 있다. 따라서, 실험 Ⅰ에서는 한 개의 STNN이 같은 단어의 변하는 발음을 모두 포함하여야 하는 부리가 따른다. 이를 극복하기 위하여, 실험 Ⅱ에서는, 각 단어를 인식하기 위한 STNN의 갯수를 1개에서 5개로 늘려서, 같은 단어가 여러 가지로 달리 발음되어도, 충분히 그 다른 발음을 별도로 구별된 STNN에서 소화할 수 있게 하였다.

확장된 network은 총 50개의 STNN들로 구성되었다. 가중치를 초기화 할때는, 같은 단어를 나타내는 5개의 STNN들을 모두 동일하게 초기화 하였고, 학습할 때는, 전체 50개의 STNN들 중에서 최종 출력값이 가장 큰 STNN만을 학습하게 된다. 또한, 인식할 때에도, 전체 50개의 STNN들 중에서 최종 출력값이 가장 큰 STNN에 해당되는 단어를 인식하게 하였다. 그림 5는 같은 단어에 대한 여러 가지 발음의 특징값들을 포함하는 확장된 STNN의 구조를 보여 주고 있다. 확장된 STNN이 포함된 인식시스템의 전체 구조가 그림 6에 설명되어져 있다.

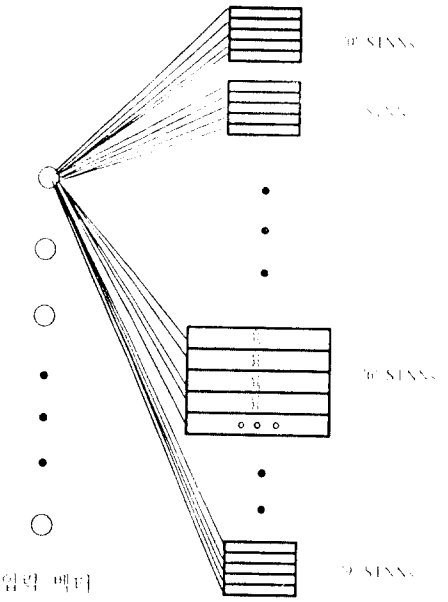

그림 5. 확장된 STNN의 구조

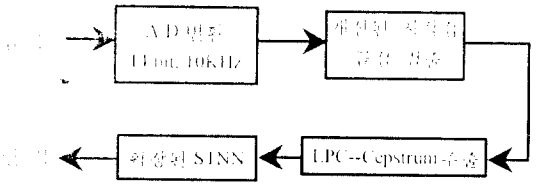

그림 6. 실험 Ⅱ에서 구성한 인식시스템

표 4. 실험 Ⅱ의 인식률

	0	1	2	3	4	5	6	7	8	9	error 수	인식률
0	61									9	9	87.1%
1		70										100%
2			70									100%
3				67	3						3	95.7%
4					70							100%
5						69				1	1	98.5%
6							70					100%
7		1						69			1	98.5%
8									70			100%
9										70		100%
총 Error 수/총 인식률											14	98.0%

표 4는 실험 Ⅱ의 결과를 보여 주고 있다. 각 단어를 나타내는 STNN의 갯수를 1개에서 5개로 늘림으로, 실험 Ⅰ에서 나타난 0과 9, 5와 9, 그리고, 1과 7 사이의 오류를 감소시켰으며, 각 숫자음의 인식률을 98%로 증가시켰다. 그리고, 7자리 연속음의 인식률을 89%로 향상시켰다. 표 5는 실험 Ⅰ과 실험 Ⅱ에 의한 각 단어에 대한 인식률의 향상을, 표 6은 실험 Ⅰ과 실험 Ⅱ에 의한 연속음의 인식률 향상을 보여 주고 있다.

표 5. 초기 7자리 연속음, 실험 Ⅰ, 그리고 실험 Ⅱ의 각 숫자음에 대한 인식률

	인 식 률
초기 7자리 연속음	82.0%
실험 Ⅰ	93.0%
실험 Ⅱ	98.0%

표 6. 초기, 실험 Ⅰ, 그리고 실험 Ⅱ의 일곱 연속 숫자음에 대한 인식률

	인 식 률
초기 7자리 연속음	28%
실험 Ⅰ	61%
실험 Ⅱ	89%

Ⅳ. 결 론

본 논문에서는, 7자리 연속 숫자음의 인식을 시도하였으며, 초기의 인식률은 28%로 저조함을 보였다. 인식률의 향상을 위하여, 한국어 단어의 발음 특성에 착안하여, 시작점은 에너지값과 zero-crossing rate의 비교로써 검출하고, 끝점은 에너지값 만으로 검출하였다. 이는 또한, STNN의 미비점을 보완하는 역활도 포함한다. 이 경우, 7자리의 연속음이 모두 맞았을 때의 인식률은 61%로 향상되었다. 따라서, 연속 단어의 인식에서, 음의 앞부분과 음의 끝부분을 검출하는 방법을 서로 달리함으로 보다 정확한 단어 사이의 경계를 설정할 수 있음을 보였다. 또한, 같은 단어일지라도, 발음이 여러 가지로 변하는 경우를 생각하여 각 단어를 나타내는 STNN의 갯수를 1개에서 5개로 늘려서, 같은 단어가 여러 가지로 달리 발음되어도, 충분히 그 다른 발음을 별도로 구별된 STNN에서 소화할 수 있게 하였다. 이로써, 7자리의 연속음이 모두 맞았을 때의 인식률을 89%로 향상시켰다. 이는 STNN이 같은 음이라도 다르게 발음되는 연속음의 특성을 잘 나타낼 수 있는 효율적인 모델임을 보여준 것이다.

본 논문에서 시도한 시작점과 끝점검출의 성능은, 좀 더 구체적으로 구분된 한국어 발음 특성에 기초하여 시도한다면, 더 나은 효과를 얻을수 있을 것으로 기대한다. 또한, 본 논문에서는 각 숫자음에 대하여 일률적으로 5개의 STNN을 모델링 하였으나, 각 숫자음에 대하여 최적의 갯수로 모델링한다면 더욱 인식률을 높이거나, 또는 전체 시스템의 복잡도를 줄일 수 있을 것으로 생각된다.

향후 연구 과제로서는, 본 논문에서 발전시킨 시스템과 기존의 MLP나 TDNN 또는 HMM 인식기와의 성능 비교를 들 수 있다. 또한, 본 연구에서는 화자종속 시스템에 촛점을 맞추었으나, 화자독립 시스템의 발전도 앞으로의 연구과제로 생각할 수 있다.

참 고 문 헌

1. P. Demichelis, "On the Use of Neural Networkks for Speaker Independent Isolated Word Recognition," *Int. Conf. on Acoust., Speech, and Signal Proc.*, May 1989.
2. 이종석, 이상욱, "신경망과 구문 분석을 이용한 한국어 연결 숫자음 인식," 대한 전자공학회 논문지, pp. 21-30, 1993.
3. 이영호, 정홍, "음절을 기반으로한 한국어 음성인식," 대한 전자공학회 논문지, pp. 11-22, 1994.
4. R. Cshalkof, *Pattern Recognition, Statistical, Structural and Neural Approaches*, Jone Wiley & Sons Inc., pp. 194-195, 1992.
5. R. Hetch-Nielson, *Neuro Computing*, Addison Wesley, 1990.
6. J. A. Freeman and D. M. Skapura, *Neural Networks*, Addison Wesley, 1990.
7. 백승우, 홍승홍, "시공간 패턴인식 신경회로망을 이용한 격리단어의 인식," 인하대학교 석사 졸업논문, 1993.
8. J. Zurada, *Artificial Neural Systems*, West Info Access, 1992.
9. 이종석, 정재호, "시공간 신경회로망을 이용한 한

국어 숫자음 인식," 한국통신학회지, pp. 771-779, March, 1995.

10. J. S. Lee and H. Chung, "Recognition of Digits Using Spatio-Temporal Neural Network," SPIE's International Symposium on Applications and Science of Artificial Neural Networks, April, 1995.
11. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, pp. 396-452, 1978.
12. L. R. Rabiner and M. R. Samber, "An Algorithm for Determining the Endpoints of Isolated Utterances," *Bell Tech. Journal*, vol. 54, no. 2, pp. 297-315, Feb. 1975.
13. P. Strobach, *Linear Prediction Theory, a Mathematical Basis for Adaptive System*, Springer-Verlag, pp. 13-36, 1990.

이 종 식(Jong Sik Lee) 학생회원

한국통신학회 제20권 제3호 참조

정 재 호(Jae Ho Chung) 정회원

1982년 : 美國 University of Maryland (공학사)

1984년 : 美國 University of Maryland (공학석사)

1990년 : Georgia Institute of Technology (공학박사)

1984년~1985년 : 美國 국방성산하 해군연구소, 신호처리실, Electronic Engineer

1991년~1992년 : 美國 AT&T Bell 연구소, 음성 신호처리 연구실, 연구원 (Member of Technical Staff)

1992년~현재 : 인하대학교 전자공학과, (현)부교수

1995년~현재 : 한국전자통신연구소, 자연어처리연구실 초빙 연구원