

# 시퀀스-피이드백 신경회로망을 이용한 한국어 숫자음 인식

## Korean Digits Recognition using Sequence-feedback Neural Network (SFNN)

최정철, 이기영\*, 최종환\*\*, 최갑석\*\*\*

(Jung Chul Choi, Ki Young Lee\*, Jong Hwan Choi\*\*, Kap Seok Choi\*\*\*)

### 요약

본 논문에서는 음성신호의 시간정보와 상관성을 포함할 수 있는 방법으로 시퀀스-피이드백 신경회로망(sequence-feedback neural network : SFNN)을 제안하였으며, 그에 의한 단어인식을 수행하였다. 이 방법에서는 음성을 인식할 때 순차적으로 과거의 출력을 네트워크 입력으로 피이드백시켜 신경회로망으로 하여금 음성의 상관성과 시간정보를 갖도록 하는 순차적인 피이드백 순환구조의 시퀀스-피이드백 신경회로망을 이용하였으며 시퀀스-피이드백 신경회로망의 입출력층의 수를 줄이는 방법으로 등분 분할을 사용하였다.

본 논문에서 제안한 방법으로 한국어 숫자음을 대상으로 실험한 결과 92.5%의 인식율을 얻었으며, 피이드포워드형 신경회로망에서의 오인식율을 1/3배 정도로 감소시켰다.

### Abstract

This paper proposes sequence-feedback neural network(SFNN) as a word recognition method which includes time-information and correlation of speech signal. To be time-informed and correlated with speech, this sequence-feedback neural network consist of sequential feedback recurrent architecture. Equi-segmentation is used to reduce the number of input layers of sequence-feedback neural network.

Using the proposed method, the recognition rate of 92.5% is acquired for Korean digits, and the inaccurate recognition is reduced to 1/3 times.

---

명지대학교 전자공학과

Department of Electronic Engineering, Myong-Ji University

\*관동대학교 전자통신공학과

Department of Electr. Comm. Engineering, Kwan Dong University

\*\*충남전문대 전자과

Department of Electronics, Chung-Nam Junior College

\*\*\*명지대학교 정보통신공학과

Department of Inf. Comm. Engineering, Myong-Ji University

접수일자 : 1992년 11월 20일

## I. 서 론

음성은 인간 사이에서의 자연스러운 통신 수단이지만, 인간과 기계 사이의 통신 수단으로 이용되지 못하고 있는 것은 음성신호가 불규칙신호이며, 발생 시간의 변동과 주파수축의 변동특성을 지니고 있기 때문이다<sup>[1,2]</sup>. 이러한 문제를 해결하기 위해 여러 인식방법들이 제안되어 왔으며 그인식방법들은 인식방법에 따라 정합에 의한 방법<sup>[3-6]</sup>과 통계적인 방법<sup>[7,8]</sup>, 신경회로망에 의한 방법<sup>[9-18]</sup>으로 크게 구분될 수 있다.

여기서 최근 활발히 연구되어 있는 신경회로망은 인간과 사고와 판단특성을 모델링한 정보처리 방법으로 접근방법에 따라 정적인 접근방식과 능동적인 접근방식으로 나누어 지고 있다<sup>[9]</sup>. 정적인 접근방식에서는 수동적이든 자동적이든 먼저 세그멘테이션을 한후 그것으로부터 특징을 추출하여 신경회로망에 입력하므로써 미지의 음성을 인식한다. 그러나 음성은 비선형적인 시간축상에 존재하기 때문에 음성신호의 어떤 부위를 세그멘테이션하느냐에 따라 인식 성능이 좌우된다는 점<sup>[10,11]</sup>과 대표적인 특징을 어떻게 추출하느냐에 따라 의존성이 커진다는 점<sup>[12,13]</sup>등이 문제점으로 부각되고 있다.

이러한 정적인 접근방식의 문제점을 해결하기 위한 방법으로 연구되고 있는 신경회로망의 동적인 접근방식은 음성신호의 시변특성에 상관없이 시불변적으로 음성신호의 특징을 파악할 수 있도록 신경회로망을 재구성한 방식으로 TDNN<sup>[14,15]</sup>과 Recurrent Net<sup>[16-18]</sup>등을 들 수 있다. 여기서 TDNN은 시간지연에 따라 중첩형태로 신경회로망을 적용하여 시불변적인 개념을 추가한 방법으로 제시되었으나, 음성신호의 시간축 소단위 입력을 확장된 신경회로망을 이용하기 때문에 복잡한 알고리즘을 요구한다. 또한, Recurrent Net를 이용한 단어인식에서는 순환구조형의 구조에 의해 시간개념을 포함하는 방법으로 미지단어를 인식하였으나, 음성신호의 현재와 과거의 특징을 모두 동일한 신경회로망에 수렴시키고자 하는 학습방법의 한계때문에 정적인 접근방식보다도 낮은 90% 미만의 인식율을 얻었다<sup>[9]</sup>. 이를 개선하기 위해 개발된 Seiichi Nakagawa<sup>[18]</sup>의 시퀀셜 신경회로망에서는 Recurrent Net와 DTW를 이용하여 비교적 높은 인식결과를 얻고 있으나 DTW 알고리즘을 적용하기 때문에 연산량이 많아지는 문제점이 있다.

이러한 문제점들을 보완하기 위하여 본 연구에서는 Recurrent Net 피이드백 구조가 음성의 특징 시계열로 하여금 상관성과 시간정보를 갖도록 한다는 점과 음성신호의 현재와 과거의 특징들을 서로 다른 신경회로망에 연결하므로써 정적인 패턴인식에 뛰어난 신경회로망의 특성을 충분히 이용한다는 점에서 착안하여 현재의 신경회로망이 현재의 특징과 함께 과거의 특징을 포함한 신경회로망의 출력을 얻을 수 있는 시퀀스-피이드백 신경회로망(sequence-feedback neural network : SFNN)을 제안하고, 이것을 이용한 단어인식 방법을 제시하고자 한다. 본 연구에서 제시한 방법의 유효성을 확인하기 위하여 정적인 접근방식인 피이드포워드형 신경회로망에 의한 단어인식 방법과 비교·검토하였다.

## II. 시퀀스-피이드백 신경회로망을 이용한 음성인식

피이드포워드형의 신경회로망을 이용한 인식방법은 정적인 패턴에 인식방법으로는 뛰어나지만, TDNN이나 시퀀셜 신경회로망들은 학습과정 및 학습방법에서 매우 복잡한 알고리즘을 요구하는 문제점이 있다. 따라서 본 연구에서는 신경회로망의 구조가 음성신호와 같은 시계열패턴에 있어서 상관성이나 시간정보를 많이 포함하도록 하기 위해 새로운 신경회로망의 구조인 시퀀스-피이드백 신경회로망(sequence-feedback neural network : S-FNN)을 제안하였다. 그림1에서 시퀀스-피이드백 신경회로망의 구조와 이를 이용한 음성인식시스템의 구성도를 보이고 있다.

이 시퀀스-피이드백 신경회로망은 신경회로망의 출력을 순차적으로 다음 신경회로망의 입력에 피이드백 시켜줌으로써 음성신호의 상관성과 시간정보를 갖도록 하였다. 먼저 음성이 입력되면 LPC 분석하여 특징벡터 시계열을 추출하고, 시퀀스-피이드백 신경회로망에 입력시키기 위하여 전 음성신호의 특징벡터 M개 프레임의 시계열을 N개의 등분 구간으로 분할하여 각 구간의 프레임 수를 M/N으로 하며 k번째 구간의 입력벡터를 다음식과 같이 구성한다.

$$\bar{a}_k = \text{LPC} \left[ \frac{N}{M} \sum_{i=0}^{M/N-1} R_i \right] \quad (1 \leq k \leq N) \quad (1)$$

여기서,  $R_i$ 는 i번째 프레임의 자기상관계수벡터,  $\text{LPC}[\cdot]$ 는 자기상관계수벡터로부터 선형예측분석,

$a_k$ 는  $k$ 번째 구간의 선형예측계수벡터,  
 $\frac{M}{N}$ 는 각 구간의 프레임 수이다.

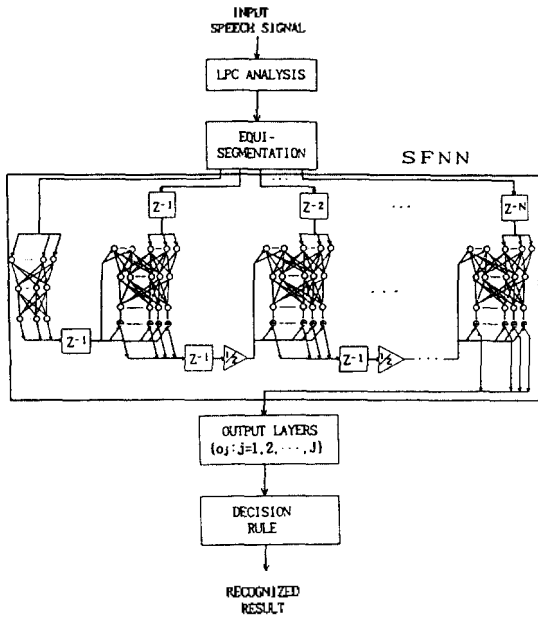


그림 1. 시퀀스-피이드백 신경회로망을 이용한 음성인식 시스템

그림 Fig 1 Speech recognition system using the sequence feedback neural network(SFNN)

2-1 학습과정

시퀀스-피이드백 신경회로망에 의한 음성인식에서 음성의 상관성과 시간개념을 포함하기 위하여 다음과 같은 학습과정을 수행한다.

- (i) 학습단어를 LPC 분석한 후 전 LPC 특징벡터 시계열을 N구간으로 등분하고, 식 (1)에 의해 각 구간의 학습용 입력벡터를 작성한다.
- (ii) 제1구간의 신경회로망은 각 학습단어의 제1구간에 해당하는 입력벡터를 입력층으로 하여 오차역전파법에 의해 학습한다.
- (iii) 제2구간의 신경회로망은 (ii)과정에서 학습된 출력층을 피이드백하여 제2구간의 입력벡터와 함께 입력층을 구성하고 오차역전파법으로 학습하며, 구간번호  $k=3$ 으로 한다.
- (iv) 제 $k$ 구간의 신경회로망은 제  $k-1$ 구간에서 학습된 출력층과 제  $k-2$  구간의 출력층의 합에  $1/2$ 을 곱한후 피이드백하여 context층  $C(k)$ 를 구성하고 입력벡터  $\bar{a}_k$ 와 함께 입력층을 구성하며

오차역전파법에 의해 학습한다. 이때, context 층을 구성하는 식은 다음과 같다.

$$C(k) = 1/2(O_{k-1} + O_{k-2}) \quad (2)$$

여가서,  $C(k)$ 는 제 $k$ 구간 신경회로망의 context 층,  $O_{k-1}$ 과  $O_{k-2}$ 는 각각 제  $k-1$ 구간과 제  $k-2$ 구간 신경회로망의 학습된 출력층이다.

(v)  $k=k+1$ 하여  $k=N+1$  이면 학습과정을 종료하고 아니면 과정(iv)부터 반복한다.

2-2 인식과정

2-1절의 학습과정에서 시퀀스-피이드백 신경회로망이 학습되면 미지단어를 입력하여 단어를 인식한다. 즉, 미지단어가 입력되면 먼저 LPC분석 한후, N개의 등간격으로 분할하며, 이전 구간의 출력층들이 피이드백되어 식(2)에 의해 작성된 context층  $C(k)$ 가 각 구간의 입력벡터  $\bar{a}_k$ 와 함께 신경회로망에 입력되므로써 미지단어를 인식한다.

이 인식과정은 다음과 같으며 그림1과 같이 시퀀스-피이드백 신경회로망 음성인식 시스템으로 나타내었다.

- (i) 입력단어를 LPC분석한 후 LPC특징벡터 기계열을 N구간으로 등분하고, 식(1)에 의해 각 구간의 입력벡터를 작성한다.
- (ii) 각 구간의 입력벡터를 해당하는 구간의 시퀀스-피이드백 신경회로망에 입력한다.
- (iii) N번째 구간의 시퀀스-피이드백 신경회로망으로부터 얻은 출력층에서 최대 출력층에서 최대 출력값으로 갖는 유니트를 선택하여 입력된 미지단어를 인식한다.

인식과정 (iii)에서 시퀀스-피이드백 신경회로망은 각 구간의 출력층이 순차적으로 피이드백되어 이웃한 신경회로망에 이전 구간의 출력층들이 누적되면서 context층을 구성하기 때문에 N번째 구간 시퀀스-피이드백 신경회로망의 입력층으로 구성되는 context층은 이전 구간의 모든 특성을 포함하게 된다. 따라서 그림1에서 입력된 미지단어를 인식하기 위한 결정규칙으로 N번째 시퀀스-피이드백 신경회로망의 출력층  $\{o_j; j=1, 2, \dots, J\}$  중에서 최대인 출력값의 유니트  $o_j$ 를 선택하여 미지단어를 인식한다. 이 결정규칙은 다음 식과 같다.

$$j^* = \operatorname{argmax}_j \quad (3)$$

$$1 \leq j \leq J$$

### III. 실험 및 결과고찰

#### 3-1 실험데이터

본 연구에서 사용된 데이터는 20대 남성 두 화자가 한국어 숫자음을 각각 5회씩 발성한 총 100개의 음성으로써 음성 데이터 베이스에서 학습에 사용될 학습 데이터는 두 화자가 3회 발음한 60개 데이터로 학습하였으며 인식에 사용될 시험 데이터는 나머지 40개의 데이터로 구성하였다. 이러한 음성들은 차단 주파수가 3.4[KHz]인 저역통과사킨후 샘플링 주파수가 10[KHz]인 12-bit 분해도의 A/D 변환기로 샘플링하여 저장하였다. 음성 신호의 프레임 구간은 20.0[ms] (200샘플)로 하였으며, 이동구간은 10.0[ms](100샘플)로 하여 50%가 중첩되게 하였다. 데이터 베이스의 구성을 위한 해석조건은 표4-1에 나타내었다.

표 1. 해석 조건  
Table 1. Analysis conditions.

Sampling Frequency	10	[kHz]
Cutoff Frequency of LPE	3.4	[kHz]
A/D Resolution	12	[bits]
Analysis Window	20.0	[ms]
Overlapping Interval	10.0	[ms]
Window Function	Hamming window	
LPC Order	14	

또한, 전 음성구간의 각 프레임을 LPC 분석하여 특징벡터의 시계열을 추출한 후 동분분할을 행할 때 등분된 각 구간은 음성의 특징들을 충분히 포함하고 있어야 한다. 본 실험에서 실험대상으로 하는 음성은 한음절로 구성된 한국어 숫자음이기 때문에 다음 표2와 같은 음운구조를 생성한다. 따라서 본 실험에서 동분분할을 행할 때 구간의 수 N은 음운구조를 충분히 포함할 수 있도록 N=4로 하여 수행하였다.

#### 3-2 인식결과 및 고찰

본 연구에서 제안한 시퀀스-피이드백 신경회로망을 이용한 단어인식 방법의 유효성을 확인하기 위하여 피이드포워드형 신경회로망을 이용한 단어 인식결과와 비교·검토 하였다.

표 2. 음운생성구조

Table 2. Phoneme production structures

숫자음	음운구조
零	SC-CV-VC-CS
일	SV-VC-CS
이	SV-VS
삼	SC-CV-VC-CS
사	SC-CV-VS
오	SV-VS
육	SV-VC-CS
칠	SC-CV-VC-CS
팔	SC-CV-VC-CS
구	SC-CV VS

S: 무음 C: 자음 V: 모음

#### 3-2-1 인식결과

##### (1) 피이드포워드형 신경회로망에 의한 인식결과 및 고찰

본 연구에서 사용한 피이드포워드형 신경회로망은 시퀀스-피이드백 신경회로망이 동분 분할을 수행한

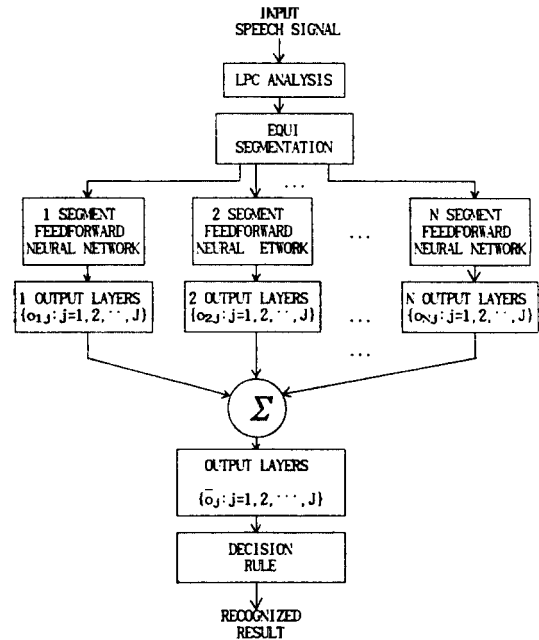


그림 2. 피이드포워드형 신경회로망에 의한 단어인식시스템

Fig 2. Feedforward neural network for word recognition system

후 분할된 각 구간이 입력층으로 적용되는 것이므로 비교실험될 피이드포워드형 신경회로망에 의한 인식 실험도 등분 분할에 의해 분할한 후 각 구간별로 피이드포워드형 신경회로망의 입력층으로 적용되도록 하였다. 이와 같이 비교실험을 수행하기 위하여 그림 2와 같이 구간별로 피이드포워드형 신경회로망에 의한 단어인식시스템을 구성하였다.

각 구간별 피이드포워드형 신경회로망 입력층의 유니트 수를 14, 중간층의 유니트 수를 12, 출력층의 유니트 수를 10으로한 3층 신경회로망을 이용하였으며, 오차역전파법으로 학습시켰다. 이때 오차역전파법의 학습파라메타는 학습률이 0.7, 관성률은 0.9로 하였으며, 오차한계는 0.001로 하였다. 학습된 후 미지의 음성 패턴에 대하여 각 구간별 신경회로망의 출력 유니트들을 다음 식과 같이 각 유니트별로 평균하여 식(3)의 최대치 선택에 의한 결정규칙에 의해 미지단어를 인식하였다.

$$\bar{o}_j = \frac{1}{N} \sum_{i=1}^N o_{ij} \quad (4)$$

여기서,  $o_{ij}$ 는 제i구간의 j번째 유니트이고,  $\bar{o}_j$ 는 단어인식을 위한 출력층의 j번째 유니트이다.

그림2의 피이드포워드형 신경회로망에 의한 단어인식을 수행하기 전에 먼저 각 구간별로 최대치를 갖는 유니트를 선택하여 표3에서 표6까지 구간별 인식 결과를 나타내었다. 표3은 피이드포워드형 신경회로망을 이용하여 얻은 제1구간의 인식 결과로 그 인식율은 80%이며, 표4는 피이드포워드형 신경회로망을 이용한 제2구간의 인식 결과로 그 인식율은 85%이다. 또한, 표5는 피이드포워드형 신경회로망을 이용한 제4구간의 인식 결과로 그 인식율은 75%이며, 표6는 피이드포워드형 신경회로망을 이용한 제4구간의 인식 결과로 그 인식율은 70%이다.

표3에서 표6까지 인식결과를 나타낸 피이드포워드형 신경회로망에 의한 인식방법에서 유사단어인 "이", "일"과 "철", "삼"과 "사", "공"과 "구"사이에서 주로 오인식이 발생하였다. 이것은 각 구간의 모음이 동일하거나 "공"의 "구"의 "우"모음과 같이 모두 후설모음이므로 스펙트럼 특성이 유사하기 때문인 것으로 보인다.

이상으로부터 피이드포워드형 신경회로망을 이용한 구간별 인식율의 평균은 약 77.5%(80+85+75+70)/4)였으나, 피이드포워드형 신경회로망에 의한

표 3. 피이드포워드형 신경회로망을 이용한 제1구간의 인식 결과

Table 3. Recognized results for the first segment using feedforward neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		3	1				1			
2			3							
3				4						
4					3					
5						4				
6					1		2			2
7		1						3		
8								1	4	
9							1			2

표 4. 피이드포워드형 신경회로망을 이용한 제2구간의 인식 결과

Table 4. Recognized results for the second segment using feedforward neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		2								
2		1	4							
3				4						
4					1					
5						4				
6					3		3			
7		1						4		
8									4	
9							1			4

표 5. 피이드포워드형 신경회로망을 이용한 제3구간의 인식 결과

Table 5. Recognized results for the third segment using feedforward neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4			1						1
1		2	1							
2			3					1		
3				2						
4				1	4					
5						2				
6						2	4			1
7		2						3		
8									4	
9										2

표 6. 피이드포워드형 신경회로망을 이용한 제4구간의 인식 결과

Table 6. Recognized results for the fourth segment using feedforward neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		2								
2			4							
3				4						
4					4					
5										
6						4				1
7								3		
8		2						1	4	
9							4			3

표 7. 피이드포워드형 신경회로망에 의한 단어인식 결과

Table 7. Recognized results for word recognition using feedforward neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									1
1		3	1							
2			3							
3				3						
4				1	4					
5						4				
6							3			
7		1						3		
8								1	4	
9							1			3

단어인식을 수행한 경우, 즉, 단어인식을 위하여 식(4)에 의해 얻은 출력층으로부터 최대치 선택에 의한 단어인식을 수행한 결과 85%의 인식율을 얻었으며, 표7에 피이드포워드형 신경회로망에 의한 단어인식 결과를 나타내었다. 이와 같이 단어인식과정에서 구간별인식과정에서 보다 높은 인식 결과를 얻을 수 있었던 것은 구간별인식과정의 제2구간과 제3구간에서 주로 오인식이 발생하였던 “일”과 “칠”, “사”와 “삼” 및 “육”과 “이” 등이 식(4)의 전 구간의 평균을 취함으로써 최대인 출력값의 유니트를 선택하는 식(3)의 결정규칙에 의해 단어를 인식하는데 기여한 것으로 생각된다.

(2) 시퀀스-피이드백 신경회로망에 의한 인식결과 및 고찰

시퀀스-피이드백 신경회로망을 이용한 음성인식은 그림1의 제시한 음성인식 시스템에 의해 수행된다. 그런데 피이드포워드형 신경회로망에 의한 인식결과와 비교하기 위하여 각 구간의 시퀀스-피이드백 신경회로망에서의 출력층에 식(3)의 결정규칙을 적용하여 각 구간별 인식결과를 나타내었다. 여기서 시퀀스-피이드백 신경회로망을 이용한 제1구간의 인식결과는 피이드포워드형 신경회로망을 이용한 제1구간의 인식결과인 표3과 같기 때문에 반복하여 제시하지 않는다. 시퀀스-피이드백 신경회로망을 이용한 제2구간에서는 피이드백된 제1구간 신경회로망의 출력층을

피이드백한 것과 음성의 제2구간에서 작성된 입력벡터와 함께 입력층을 구성하여 인식실험이 수행되며, 그 결과를 표8에 나타내었다. 이 표8에서는 피이드포워드형 신경회로망을 제2구간에 적용하여 얻었던 결과인 표4에서 유사단어 "일"과 "아" 사이에 나타났던 오인식이 줄어들었으며 90%인 인식율을 얻었다. 시퀀스-피이드백 신경회로망을 이용한 제3구간에서는 제1구간의 출력층과 이웃한 제2구간의 출력층의 합에 1/2를 곱하여 context층을 구성하고, 음성의 제3구간에서 작성된 입력벡터와 함께 입력층을 구성하여 인식실험이 수행되며, 그 결과를 표9에 나타내었다. 표9에서는 피이드포워드형 신경회로망을 제3구간에 적용하여 얻었던 결과인 표5에서 유사단어 "일"과 "이", "삼"과 "사", "공"과 "구" 사이에 나타났던 오인식이 줄어들었으며, 95%의 인식율을 얻었다. 시퀀스-피이드백 신경회로망을 이용한 제4구간에서는 제2구간의 인식 출력층과 이웃한 제3구간의 인식 출력층의 합에 1/2를 곱하면 context층을 구성하고, 제4구간에서 작성된 입력벡터와 함께 입력층을 구성하여 실험한 인식실험이 수행되며, 그 결과는 표10과 같으며, 인식율은 92.5%였다.

이상의 결과로부터 각 구간별 인식율은 피이드포워드형 신경회로망보다 순차적으로 출력층을 피이드백하여 context층을 구성하는 시퀀스-피이드백 신경회로망에 의해 향상되었음을 알 수 있었다.

표 8. 시퀀스-피이드백 신경회로망을 이용한 제2구간의 인식 결과

Table 8. Recognized results for the second segment using sequence-feedback neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		3				1				
2			4							
3				4						
4					4					
5						4				
6							2			1
7		1						4		
8									4	
9						1				3

표 9. 시퀀스-피이드백 신경회로망을 이용한 제3구간의 인식 결과

Table 9. Recognized results for the third segment using sequence-feedback neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		4					1			
2			4							
3				4						
4					4					
5						4				
6							2			
7								4		
8									4	
9							1			4

표 10. 시퀀스-피이드백 신경회로망을 이용한 제4구간의 인식 결과

Table 10. Recognized results for the fourth segment using sequence-feedback neural network

Test Ref.	0	1	2	3	4	5	6	7	8	9
0	4									
1		4								
2			4							
3				4						
4					4					
5						4				
6							4	1		1
7								3	1	
8									3	
9										3

3-2-2 인식율의 비교 및 고찰

피이드포워드형 신경회로망에 의한 구간별 인식결과인 표3에서 표6까지의 결과로부터 각 구간에 나타난 유사단어 "일"과 "이", "일"과 "칠", "삼"과 "사", "공"과 "구" 사이에 발생한 오인식 결과를 시퀀스-피이드백 신경회로망에 의한 구간별 인식결과인 표8에서 표10까지 결과로부터의 각 구간에서 나타난 유사단어 사이의 오인식 결과와 비교하기 위하여 유사단어 사이의 구간별 오인식율을 그림3과 같이 비교하였다.

그림3에서 그림4의 피이드포워드형 신경회로망에 의한 오인식은 제3구간에서 제일 크게 나타났는데 반하여 시퀀스-피이드백 신경회로망에서는 구간수가 증가할수록 오인식이 감소함을 알 수 있다. 그것은 구간 수가 증가할수록 순차적인 피이드백이 누적되어 입력층에 상관성 정보가 많이 포함되기 때문인 것으로 생각된다.

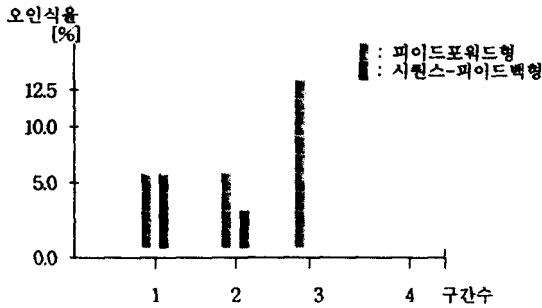
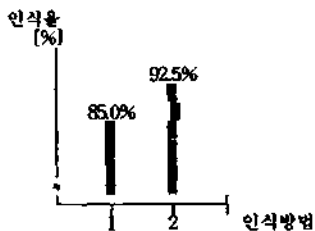


그림 3. 유사단어 사이의 구간별 오인식율의 비교  
 Fig 3. A comparison of error recognition rate for each segment between similar words

그림3에서 피이드포워드형 신경회로망에서 나타난 총 오인식율은 22.5%이며, 시퀀스-피이드백 신경회로망에 의해 총 오인식율은 7.5%로 나타났다. 이로부터 시퀀스-피이드백 신경회로망이 피이드포워드형 신경회로망보다 오인식율이 약 1/3배 정도로 감소시킬 수 있음을 알 수 있다.

또한 그림4에서 표7의 피이드포워드형 신경회로망의 단어인식결과와 표10의 시퀀스-피이드백 신경회로망의 단어인식결과로부터 얻은 인식율을 나타내었다. 여기서 피이드포워드형 신경회로망에 의해 85.0%의 인식율을 얻었으며, 시퀀스-피이드백 신경회로망에 의해 92.5%의 인식율을 얻었다.



1 : 피이드포워드형 신경회로망에 의한 단어인식  
 2 : 시퀀스-피이드백 신경회로망에 의한 단어인식

그림 4. 인식율의 비교  
 Fig 4. Comparison of recognition rate

이상에서 볼 수 있는 바와 같이 시퀀스-피이드백 신경회로망을 이용한 단어인식에서 피이드포워드형 신경회로망을 이용한 경우 보다 높은 인식율을 얻을 수 있었던 것은 각 구간의 신경회로망에 의한 출력층이 순차적으로 피이드백되어 단어가 가진 특성들이 누적되어 context층을 구성하기 때문에 입력층이 상관성정보를 포함하게 되어 단어의 인식성능이 향상되고 유사단어 사이의 오인식도 감소되었기 때문인 것으로 생각된다.

IV. 결 론

본 논문에서는 시퀀스-피이드백 신경회로망을 제안 하였으며, 한국어 숫자음을 대상으로 음성인식 실험을 하여 피이드포워드형의 신경회로망에 의한 음성인식결과와 비교 검토한 결과 다음과 같은 결론을 얻을 수 있었다.

- (1) 유사단어의 오인식율 비교결과 시퀀스-피이드백 신경회로망을 이용하였을 때의 오인식율이 피이드포워드형 신경회로망의 오인식율보다 1/3배 정도로 감소되었다.
- (2) 시퀀스-피이드백 신경회로망에 의한 음성인식에서 피이드포워드형 신경회로망에 의한 것보다 향상된 인식율을 얻었다.

따라서, 시퀀스-피이드백 신경회로망에 의한 단어 인식방법은 피이드포워드형 신경회로망에 의한 인식방법보다 우수한 음성인식방법임을 확인하였다.

참 고 문 헌

1. J.L.Flanagan, "Voices of Men and Machine," J. Acoust. Soc. Am., Vol.51, Mar.1971.
2. L.H.Witten, Principles of Computer Speech, Academic Press, 1982.
3. L.R.Rabiner and C.E.Schmidt, "Appication of Dynamic Time Warping to Connected Digital Recognition," IEEE Trans., Trans., Acoust., Speech., Signal Processing, Vol ASSP-28, No.4, pp.377-388, 1980.
4. C.S.Myers and L.R.Rabiner, "Connected Digit Recognition Using A Level-building DTW Algorithm," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-29, No.3, pp.351-363, 1981.
5. J.E.Shore, "Discrete Utterance Speech Recognition Without Time Normalization," ICASSP, 5.4, 1982.



6. K.C.Pan, F.K.Soong, L.R.Rabiner, "A Vector-quantization Based Preprocessor For Speaker-independent Isolated Word Recognition," IEEE Trans. Apeech, Signal Processing, Vol.ASSP-33, NO.3, pp.546-553, 1985.
7. A.Averbuch, et al., "Experiments With The Tangora 20,000 Word Speech Recognizer," ICASSP 87, pp. 701-704, dApril, 1987.
8. L.R.Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, Vol.77, No.2, Feb.1989.
9. Alexander Waibel, "Neural Network Approaches for Speech Recognition," Advances in Speech Signal Processing, Marcel Dekker, Inc., New York Besel Hong Kong, 1992.
10. J.L.Eلمان, D.Ziper, "Learning The Hidden Structure of Speech," Technical Report, Univ. California, San Diego, Feb. 1987.
11. R.P.Lippmann, B.Gold, "Neural Net Classifiers Usefulfor Speech Recognition," IEEE int.Conf. Neural Networks, Jun. 1987.
12. S.Peeling, R.Moore, "Experiments in Isolated Digit Recognition Using the Multi-layer Perceptron," Technical Report 4073, Royal Speech and Radar Establishment (RSRE), Dec.1987.
13. B.R.Kammerer, W.A.Kupper, "Experiment For Isolated-word Recognition With Sigle And Two Layer Perceptrons," Neural Networks, Vol.3, PP. 693-706, 1990.
14. A.Waibel, T.Hanazawa, G.Hinton, K.Shikano, K. Lang, "Phoneme Recognition Time-delay Neural Network," IEEE trans. Acoust. Speech, Signal Processing, Vol.ASSP-37, No.3, pp.328-337, 1989.
15. 南泰浩, 澤井秀文, 官武正典: "Large Vcabulary Spoken Word Recognition Using Time-Delay Neural Network Phoneme Spotting and LR-Parsing," 電子情報通信學會論文誌, J73-D-II, 6, pp.788-795, 1990.
16. R.Watrous, "Speech Recognition Using Connecti-onist Networks," Ph.D. thesis, University of Pennsylvania, 1988.
17. G.Z.Sun, H.H.Chen, Y.C.Giles, "Recurrent Neural Networks, Hidden Markov Models And Stochastic Grammars," IJCNN, Vol.I, pp.719-726, 1990.
18. 中川聖一, 早川 勳, "Speech Recognition Using Vari-ous Sequential Networks," 電子情報通信學會論文誌 D-II, Vol.J74-D-II, No.9, pp.1174-1183, 1991.

▲최 정 철

1967년 6월 25일생 :  
 1991년 2월 : 명지대학교 전자공학과 졸업  
 1993년 2월 : 명지대학교 대학원 전자공학과 석사과정 졸업(공학석사)

▲이 기 영



1961년 5월 7일생  
 1984년 2월 : 명지대학교 전자공학과 졸업  
 1986년 2월 : 명지대학교 대학원 전자공학과 석사과정 졸업(공학석사)  
 1992년 2월 : 명지대학교 대학원 전자공학과 박사과정 졸업(공학박사)  
 1993년 3월 ~ 현재 : 관동대학교 전자통신공학과 조교수

▲최 종 환



1956년 1월 8일생  
 1978년 2월 : 명지대학교 전자공학과 졸업  
 1980년 2월 : 명지대학교 대학원 전자공학과 석사과정 졸업(공학석사)  
 1992년 2월 : 명지대학교 대학원 전자공학과 박사과정 수료  
 1988년 3월 ~ 현재 : 충남전문대학 전자과 조교수

▲최 김 석

11卷3號(92년 6월 30일) 참조