

# 뉴럴-퍼지패턴매칭에 의한 단어인식에 관한 연구

## (A Study on Word Recognition Using Neural-Fuzzy Pattern Matching)

李 基 榮\*, 崔 甲 石\*\*

(Ki Young Lee and Kap Seok Choi)

### 要 約

본 연구에서는 음성의 스펙트럼시계열로 적절한 패턴을 작성하고 음성인식율을 향상시키기 위하여 퍼지패턴 매칭을 개선한 뉴럴-퍼지패턴매칭에 의한 단어인식방법을 제시하였다. 이 방법에서는 신경회로망의 연상기억에 의해 패턴을 작성하여 주파수변동을 흡수하고 퍼지패턴매칭에 의해 유사도를 측정하므로써 시간변동의 문제를 보완하였다. 또한, 음성의 특징파라미터는 2진화 스펙트럼을 사용하며, 유사도는 논리연산에 의해 측정되기 때문에 왜곡척도를 이용한 DTW 방법에 비해 기억용량과 계산량이 매우 작다. 이방법의 인식성능을 평가하기 위하여 28 개의 도시명을 대상으로 인식실험을 수행하였으며 DTW 방법, 퍼지패턴매칭방법과 비교 검토한 결과, 뉴럴-퍼지패턴매칭에 의한 인식성능이 다른 방법보다 우수함을 확인하였다.

### Abstract

This paper presents the word recognition method using a neural-fuzzy pattern matching, in order to make a proper speech pattern for a spectrum sequence and to improve a recognition rate. In this method, a frequency variation is reduced by generating binary spectrum patterns through associative memory using a neural network, and a time variation is decreased by measuring the similarity using a fuzzy pattern matching. For this method using binary spectrum patterns and logic algebraic operations to measure the similarity, memory capacity and computation requirements are far less than those of DTW using a conventional distortion measure. To show the validity of the recognition performance for this method, word recognition experiments are carried out using 28 DDD city names and compared with DTW and a fuzzy pattern matching. The results show that our presented method is more excellent in the recognition performance than the other methods.

### I. 서 론

\*正會員, 明知大學校 電子工學科  
(Dept. of Elec. Eng., MyongJi Univ.)

\*\*正會員, 明知大學校 情報通信工學科  
(Dept. of Inf. Comm. Eng., MyongJi Univ.)  
接受日字; 1992年 4月 24日

음성의 패턴인식에서 인식성능을 저하시키는 문제점으로는 음성의 시간변동과 주파수변동 등이 있다. 이러한 문제를 보완하는 방법으로 DTW 방법<sup>[2]</sup>과 멀티-템플릿<sup>[3]</sup> 등이 개발되었으나, 모두가 대량의 기억용량과 계산량을 필요로 하는 단점이 있다.

최근, 인간 두뇌의 근사모델인 신경회로망과 애매성을 해결하려는 퍼지이론이 음성인식에 적용되면서 음성의 시간변동과 주파수변동 문제를 해결하기 위한 연구가 활발히 진행되고 있다. 신경회로망을 이용한 방법으로는 동적 프로그램과 결합한 DNN,<sup>[4]</sup> 시간지연에 따라 신경회로망을 적용한 TDNN<sup>[5]</sup> 등이 있으며, 퍼지이론을 이용한 방법으로는 퍼지추론<sup>[6]</sup>이나 퍼지패턴매칭<sup>[7]</sup>에 의한 방법등이 있다.

그러나 이들 방법에서는 전형적인 대역통과필터분석, FFT 분석, 선형예측분석에 의해 추출한 특징을 입력으로 하여 인식을 수행하기 때문에 발생시간에 따른 시간변동이나 성도의 순간적인 위치변화와 유연성 때문에 생기는 주파수변동의 문제를 인식과정에서 흡수하지 않으면 안되므로, 많은 학습데이터를 필요로 하거나, 시행착오에 의한 인식방법의 개정 및 인식방법의 구조개선등이 필요하다.

본 연구에서는 이문제들을 보완하기 위하여, 스펙트럼 특징으로 구성된 패턴을 작성하는 과정에서 동일음성의 성도특성과 음원특성이 동일하게 나타날 수 있도록 신경회로망의 연상기억특성을 이용하여 주파수변동의 폭을 흡수하였으며, 비선형적인 발생시간에 따른 시간변동의 폭을 모두 포함할 수 있도록 하기 위하여 멤버쉽함수를 이용하여 표준패턴을 작성하고 퍼지패턴매칭을 이용하여 인식을 수행하였다. 따라서, 본 연구에서는 신경회로망과 퍼지패턴매칭을 결합한 뉴럴-퍼지패턴매칭(a neural-fuzzy pattern matching)에 의한 음성인식방법을 제안하고자 한다.

본 연구의 뉴럴-퍼지 패턴매칭에 의한 음성인식방법의 성능을 평가하기 위하여 28개의 도시명을 대상으로 음성인식을 수행하였으며, DTW 방법과 퍼지패턴매칭의 기억용량, 계산량 및 인식율과 비교, 검토하였다.

## II. 음성의 특징파라미터 추출

음성이 입력되면 4.0 kHz 차단주파수의 저역통과필터를 통해 10 kHz 의 샘플링주파수로 A/D 변환하며, 특징파라미터로 구성된 패턴을 작성하기 위해 선형예측분석에 의해 스펙트럼을 추정한다. 스펙트럼의 시계열에 나타나는 음성의 피크주파수는 음원특성과 함께 성도의 전달특성을 수반하기 때문에 음성인식에서 중요한 특징파라미터의 역할을 한다.

따라서 본 연구의 음성패턴을 구성하는 특징파라미터는 스펙트럼시계열의 각 피크주파수를 청각특성에 근사한 1/3 옥타브 대역으로 변환하고 2 진화 처리한 2진화 스펙트럼이다. 그림 1 에 선형예측분석에 의하여 특징파라미터를 추출하는 과정을 나타내었다. 여기에서 음성신

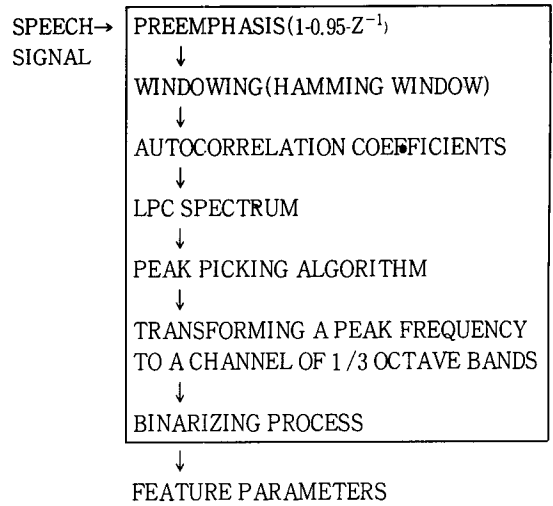


그림 1. 선형예측분석에 의한 특징파라미터의 추출과정  
Fig. 1. Procedure for extracting feature parameters using linear prediction analysis.

호가 입력되면 선형예측분석하여 스펙트럼을 추정하고 이 스펙트럼의 피크주파수를 피크피킹 알고리즘<sup>[9]</sup>에 의하여 검출한다. 한 스펙트럼의 피크주파수들을 청각특성에 근사시키기 위해 다음 식에 의해 피크주파수를 1/3 옥타브 대역<sup>[11]</sup>에 해당하는 채널로 변환한다.

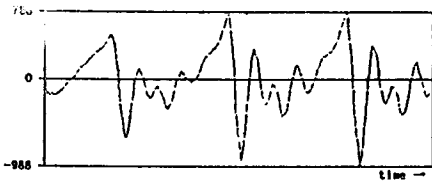
$$k_j = INT[3 \times \log_2 \frac{f_j}{f_0}] \quad (1)$$

여기서  $k_j$ 는 1/3 옥타브 대역의 18개 대역 중에서  $j$  번째 피크주파수에 해당하는 대역의 채널을 나타내며  $f_0$ 는 1 채널의 중심주파수 100Hz 이며  $f_j$ 는  $j$ 번째 피크주파수이고  $INT[ \cdot ]$  정수화 함수이다.

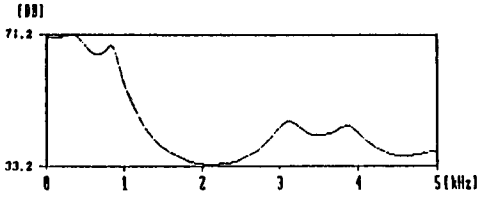
그림 2에 식(1)을 이용하여 음성의 스펙트럼을 2진화 스펙트럼으로 변환하는 예를 보이고 있다. 식(1)에 의해 피크주파수를 해당하는 1/3 옥타브 대역의 각 채널로 변환하고 그 채널을 1 로, 나머지 채널을 0 으로 표시하여 각 스펙트럼 단위로 특징파라미터를 추출하며 이와 같은 과정을 음성의 전 스펙트럼시계열에 수행한다.

## III. 신경회로망에 의한 패턴작성

음성의 동일한 음운은 이상적으로 성도의 전달특성과 음원특성이 동일해야 하지만 같은 화자일지라도 성도의 순간적인 위치 변화와 유연성 때문에 동일한 음운의 스펙트럼이 변화하여 주파수변동을 일으킨다. 따라서, 음성으



(a)



(b)

PEAK FREQUENCIES ( $f_{p_j}$ )	SELECTED CHANNEL ( $k_j$ )
$f_{p1} = 378 \text{ Hz}$	$k_1 = 5$
$f_{p2} = 986 \text{ Hz}$	$k_2 = 9$
$f_{p3} = 3194 \text{ Hz}$	$k_3 = 14$
$f_{p4} = 3916 \text{ Hz}$	$k_4 = 15$

(c)

CH	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
BS	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0

(d)

그림 2. 1/3옥타브 2진화 스펙트럼

- (a) 음성신호
- (b) LPC 스펙트럼
- (c) 피크주파수와 채널
- (d) 1/3 옥타브 2진화 스펙트럼

Fig. 2. 1/3 octave binary spectrum.

- (a) Speech signal,
- (b) LPC spectrum,
- (c) Peak frequencies and channels,
- (d) 1/3 octave binary spectrum.

로부터 추출된 1/3옥타브 2진화스펙트럼은 주파수변동의 영향을 받으며, 인식성능을 저하시키는 요인이 된다. 본연구에서는 주파수 변동에 따른 인식성능의 저하를 방지하기 위해 신경회로망 특징 중의 하나인 연상기억능력을 이용하기로 한다. 즉, 동일한 음운의 유사한 2진화스펙트럼들을 입력으로 하고 그들의 대표 2진화스펙트럼을 교사용답으로 하여 학습시키므로써 신경회로망은 입력음성의 유사한 2진화스펙트럼들을 각각의 대표 2진화스펙트럼으로 변화시켜 동일한 음운에서 발생하는 주파수변동의 폭을 감소시킨다.

패턴작성을 위한 신경회로망은 빠른 학습능력을 갖는 역전파학습(back-propagation learning) 알고리즘을 적용하였으며 이를 학습시키기 위해 교사용답으로는 종래의 벡터양자화에서 사용하는 집단화(clustering) 알고리즘을 사용하여 얻은 대표 스펙트럼을 2진화스펙트럼으로 변환하여 적용하였다. 그림 3은 신경회로망에 의해 2진화 스펙트럼패턴을 작성하는 전처리 과정이며, 그 학습과정은 다음과 같다.

- (a) 음성을 선형예측분석한 후, 집단화 알고리즘에 의해 음성의 코드북을 생성하며 각 집단의 대표 스펙트럼의 특징파라미터를 추출한다.
- (b) 각 집단의 대표 스펙트럼의 특징파라미터를 교사용답으로 하고 그에 해당하는 집단 스펙트럼들의 특징파라미터들을 학습데이터로 하여 신경회로망을 훈련시킨다.

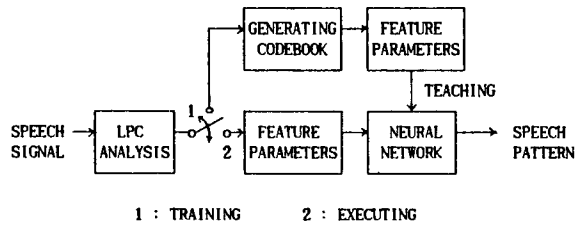


그림 3. 신경회로망에 의한 패턴작성의 구성도

Fig. 3. Block diagram for making pattern using a neural network.

이상의 학습과정에 의해 신경회로망이 훈련되면 다음과 같은 수행과정에 의해 패턴을 작성한다.

- (a) 입력된 음성으로부터 선형예측분석에 의해 특징파라미터를 추출하고 신경회로망에 입력하면 해당하는 집단의 대표 특징파라미터를 출력한다.
- (b) 위와 같은 과정을 음성의 전 시계열에 적용하여 각 집단의 특징파라미터로 구성된 패턴을 작성한다.

그림 4에 신경회로망에 의해 패턴을 작성하는 예를 보이고 있다. 여기서 (a)는 음성신호, (b)는 음성신호의 스펙트럼 시계열, (c)는 식(1)에 의한 2진화 스펙트럼 패턴, (d)는 신경회로망에 의해 작성한 2진화 스펙트럼 패턴이다.

#### IV. 뉴럴-퍼지 패턴매칭에 의한 음성인식

뉴럴-퍼지 패턴매칭에 의한 음성인식은 신경회로망의 연상기억 능력에 의해 패턴을 작성하여 주파수변동을 흡

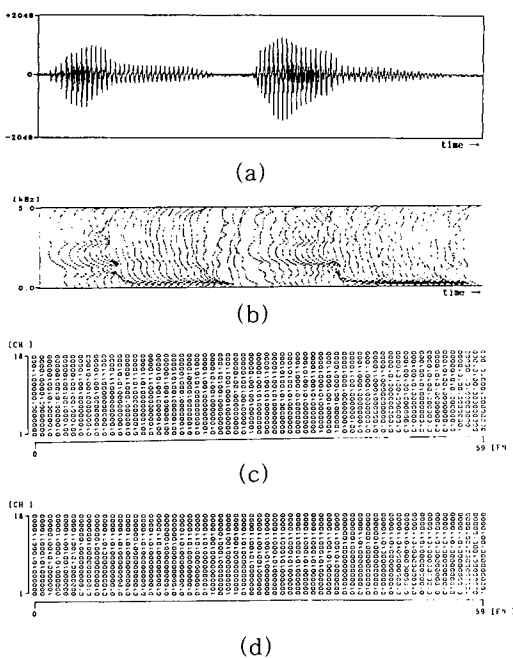


그림 4. 신경회로망에 의한 2진화 스펙트럼 패턴

- (a) 음성신호 “안산”
  - (b) 스펙트럼 시계열
  - (c) 2진화 스펙트럼 시계열
  - (d) 신경회로망에 의한 집단화 2진화 스펙트럼 시계열
- Fig. 4. Binary spectrum pattern made by a neural network .
- (a) Speech signal “ansan”,
  - (b) Spectrum sequence,
  - (c) Binary spectrum sequence,
  - (d) Binary spectrum sequence clustered by the neural network.

수하며, 퍼지패턴매칭(7)의 유사도측정을 이용하여 시간 변동의 문제를 보완한다.

뉴럴-퍼지 패턴매칭에 의한 음성인식에서 표준패턴으로 등록된 단어들을  $I = \{i_1, i_2, \dots, i_n\}$ , 단어명  $i_j$ 를 발생했을 때 특징파라미터로 구성된 패턴을  $X = \{x_1, x_2, \dots, x_n\}$ , 각 단어패턴의 귀속성을 나타내는 멤버쉽함수를  $M = \{m_1, m_2, \dots, m_n\}$  이라 하자. 여기서 멤버쉽함수  $m_j$ 는 단어패턴  $x_j$ 의 귀속등급을 나타내므로 각 단어마다 여러 번 발생하여 작성한 패턴들을 중첩하여 작성한다. 그런데 각 패턴의 주파수축의 길이는 일정하나 시간축의 길이는 서로 다르기 때문에 패턴을 중첩시키기 전에 선형신축방법<sup>[10]</sup>에 의해 길이를 정규화시킨다. 그림 5는 신경회로망에 의해 작성되는 각 단어의 패턴을 선형신축방법으로 중첩한 표준패턴을 보인다.

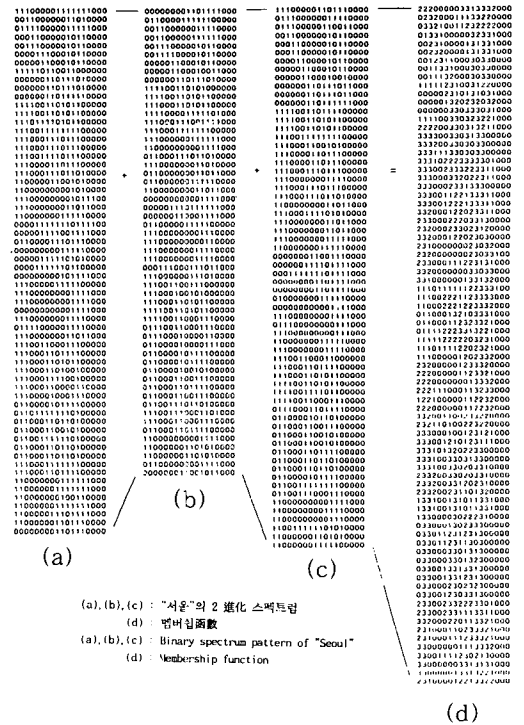


그림 5. 멤버쉽함수 (표준패턴)

- (a), (b), (c): “서울”의 2진화 스펙트럼
- (d): 멤버쉽函數

Fig. 5. Membership function (Reference pattern).

- (a), (b), (c): Binary spectrum pattern of “Seoul”,
- (d): Membership function.

그림 6은 신경회로망을 전처리 과정에서 사용하여 주파수변동을 흡수하며 표준패턴을 멤버쉽함수로 작성하여 시간변동 문제를 보완하는 뉴럴-퍼지 패턴매칭과 최대유사도결정규칙에 의해 음성을 인식하는 구성도이다. 그림 6에서 신경회로망은 IV장의 학습과정에 의해 훈련된 것이며, 음성이 입력되면 신경회로망을 거쳐 특징파라미터로 구성된 음성의 패턴을 작성한다. 또한, 같은 음성의 패턴들을 중첩한 멤버쉽함수로 표준패턴을 작성하며 시험패턴과 표준패턴 사이의 유사성을 측정하여 최대유사도결정규칙에 의해 음성을 인식한다. 시험패턴을  $y(y \in X)$ ,  $j$ 번째 등록된 단어의 표준패턴인 멤버쉽함수를  $m_j$ 라 할때 이들 사이의 유사도<sup>[7]</sup>를 측정하는 식은 다음과 같다.

$$S_y = \frac{P_{j0}}{P_{jd} - P_{j0}} \cdot \frac{P'_{j0}}{P_j - P'_{j0}} \quad (2)$$

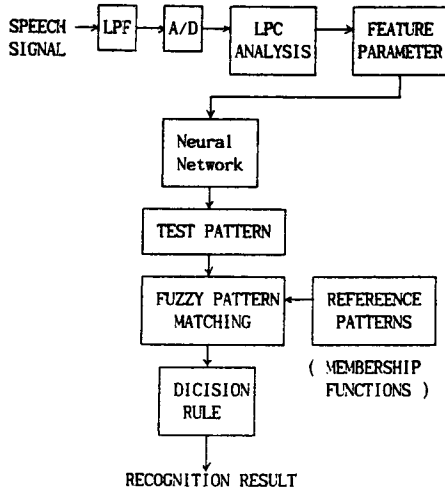


그림 6. 뉴럴-퍼지패턴매칭에 의한 음성인식 구성도  
Fig. 6. Block diagram of speech recognition using Neural-Fuzzy pattern matching .

$$\begin{aligned}
 \text{여기서, } P_{jd} &= \sum_f \sum_t m_j(f, t) & (3) \\
 P_o &= \sum_f \sum_t y(f, t) \\
 P_{jo} &= \sum_f \sum_t m_j(f, t) \cdot y(f, t) \\
 P_{j\alpha} &= \sum_f \sum_t m_j(f, t) * y(f, t)
 \end{aligned}$$

또한 미지패턴  $y$ 와 멤버쉽함수  $m_j$ 는 주파수  $f$ 와 시간  $t$ 의 2차원 패턴이며  $\cdot$ 는 두 패턴 사이의 논리곱이고  $*$ 는  $y$ 와  $m_j$ 의  $\alpha$ -cut 사이의 논리곱이다. 즉,

$$m_j(f, t) \cdot y(f, t) = \begin{cases} 1, & m_j(f, t) \geq 1 \text{ and } y(f, t) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$m_j(f, t) * y(f, t) = \begin{cases} 1, & m_j(f, t) \geq \alpha \text{ and } y(f, t) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

여기서  $\alpha$ 는 1부터 멤버쉽함수의 최대치 사이에서 선택한다.

다음 식은 결정규칙에서 최대유사도를 갖는 표준패턴의 단어명을 선택하여 미지단어를 인식하는 식이다.

$$\begin{aligned}
 j^* &= \arg \max [S_{jy}] & (6) \\
 1 &\leq j \leq n
 \end{aligned}$$

여기서,  $n$ 은 등록 단어의 수이며,  $S_{jy}$ 는 유사도 측정식이다.

V. 실험 및 고찰

1. 음성 데이터

음성 데이터는 DDD 전화 도시명 28개 (직할시 이상 6개 + 경기도내 22개(동일 지역번호의 지역명 제외))를 방음장치가 되지 않은 실험실에서 20대 남성 2인과 여성 1인이 5회씩 발성(28개×3인×5회=420개) 하여 실험데이터로 사용하였으며, 표 1과 같은 조건에 의해 음성을 분석하였다.

표 1. 분석 조건  
Table 1. Analysis conditions .

Cut-off frequency	4.0 kHz
Sampling frequency	10 kHz
A /D resolution	12bits
Window length	20 ms
LPC orders	14

뉴럴-퍼지패턴매칭에서 신경회로망을 훈련시키기 위해 각 화자가 3회씩 발성한 DDD 전화 도시명 (28개×3인×3회=252개)을 선형예측분석하여 특징파라미터를 추출하고 이들을 학습데이터로 하였으며, LBG 알고리즘<sup>[12]</sup>에 의해 코드북을 생성하였고 이 코드북의 각 스펙트럼으로부터 특징파라미터를 추출하여 교사용답으로 사용하였다. 여기서 신경회로망은 역전파 학습 알고리즘에 의해 학습되었으며, 입력층과 출력층의 유니트 수가 각각 18개, 은둔층의 유니트 수는 10개로 구성하였다.

2. 인식율의 비교

뉴럴-퍼지패턴매칭에 의한 인식성능을 검토하기 위하여, 종래의 인식방법인 DTW방법에 의한 인식결과와 신경회로망을 이용하지 않은 퍼지패턴매칭에 의한 인식결과를 비교 검토하였다.

(1) DTW 방법에 의한 인식실험

첫째 인식실험은 DTW를 이용한 인식실험이다. 여기서, 인식을 위한 DTW의 경로는 기울기가 1이고 대칭인 것으로 하였으며, DTW 방법에 의한 인식 수행시 사용한 표준패턴은 각 화자가 3번씩 발성한 음성데이터로 각 단어별 3개의 멀티템플릿으로 구성하였다. 시험데이터는 나머지 2회씩 발성한 음성데이터로 구성하여 인식실험을 수행하였으며, 그 인식결과를 표 2에 나타내었다.

(2) 퍼지패턴매칭에 의한 인식실험

신경회로망을 이용하지 않고 패턴을 작성하는 첫번째 방법은 먼저 제 II 장에 서술한 바와 같이 선형예측에 의

표 2. DTW에 의한 인식율

Table 2. Recognition rates using DTW

SPEAKER	MA	MB	FC	AVG
RATE[%]	98.2	98.2	96.4	97.6

해 추정한 스펙트럼으로부터 피크주파수와 대역폭을 모두 검출하고 1/3 옥타브 대역의 채널로 변환하여 피크주파수와 대역폭에 해당하는 채널을 1로, 나머지 채널을 0으로 표시하므로써 패턴을 작성한다. 멤버쉽함수는 이 방법으로 작성된 패턴의 3회분을 선형중첩하여 작성하며 표준패턴으로 사용한다. 따라서, 하나의 멤버쉽함수에 3개의 패턴정보가 모두 포함되어 있으므로  $\alpha$ -cut 값의 크기를 변화시키면서 인식결과를 비교할 필요가 있다. 표 3에는 피크주파수와 대역폭으로 패턴을 작성하는 경우의 인식결과를 변수로 하여 나타내었다.

표 3. 피크주파수와 대역폭의 2진화스펙트럼패턴에 대한 인식율

Table 3. Recognition rates of binary spectrum pattern with peak frequencies and bandwidths.

[%]

SPEAKER $\alpha$	1	2	3
MA	75.0	91.1	83.9
MB	85.7	92.9	73.2
FC	58.9	92.9	82.1
AVG.	73.2	92.3	79.7

표 3의 인식율에서  $\alpha$ 값이 2일때 평균 92.3%의 인식율을 얻었으며 오인식은 조음위치가 동일한 모음을 공통적으로 포함한 도시명 「부산,문산」, 「안산,안성」, 「이천,인천」 사이에서 주로 발생하였다. 이러한 오인식이 발생한 것은 1/3 옥타브 대역이 저대역 주파수에서는 좁고 고대역 주파수로 갈 수록 넓어지기 때문에 저대역 주파수 특성이 비슷한 모음을 포함한 음성은 피크주파수와 그 대역폭으로 표현된 패턴이 서로 유사하여 오인식이 발생하였다.

신경회로망을 이용하지 않은 두번째 방법에서는 첫번째 방법의 오인식을 피하기 위하여 피크주파수에 해당하는 채널만을 1로 표시하고 나머지 채널을 0으로 표시하여 패턴을 작성하고 음성인식을 수행하였다. 표 4에는 피크주파수만으로 패턴을 작성한 경우  $\alpha$ 값을 변수로 하여 인식율을 표시하였다.

표 4. 피크주파수의 2진화스펙트럼패턴에 대한 인식율

Table 4. Recognition rates of binary spectrum pattern with peak frequencies.

[%]

SPEAKER $\alpha$	1	2	3
MA	94.6	96.4	71.4
MB	89.3	92.9	62.5
FC	91.1	92.9	73.2
AVG.	92.3	94.1	69.0

표 4의 인식율에서  $\alpha$ 값이 2일때 첫번째 방법에서 발생했던 오인식을 줄일 수 있었으며, 평균 94.1%의 인식율을 얻으므로써 신경회로망을 이용하지 않은 첫번째 방법의 경우보다 인식성능을 향상시킬 수 있었다. 그러나  $\alpha$ 값이 3일때 인식율이 현저히 감소한 것은 유사도의 논리곱식(5)에서 적용될 멤버쉽함수의 폭이  $\alpha$ 값이 클수록 좁아지기 때문에 많은 오인식이 발생한 것으로 생각된다.

(3) 뉴럴-퍼지패턴매칭에 의한 인식실험

신경회로망을 이용하지 않은 인식방법에서 피크주파수만을 패턴으로 작성하는 두번째 방법에서 인식성능을 향상시킬 수 있었으므로, 본 절의 뉴럴-퍼지 패턴매칭에 의한 인식방법에서는 피크주파수만으로 표현된 특징파라미터의 시계열을 작성하며 신경회로망에 의해 전처리를 행하여 패턴을 작성하여 음성인식을 수행한다. 이 방법에서 신경회로망에 의해 사상되는 집단의 수는 음운의 수와 일치시켜야 하지만 본 연구에서는 유사한 스펙트럼들을 한 집단으로 가정하므로 단어별 집단의 수를 변화시키면서 인식율을 비교하였다.

표 5에는 각각 신경회로망에 의해 학습된 집단의 수가 4, 8, 16 일 때,  $\alpha$ 값을 변수로 한 인식율을 나타내었다. 한 단어에 대해 신경회로망에 의해 학습된 집단의 수가 4인 경우 가장 높은 인식율을 얻을 수 있었던 것은 한 단어의 음운수와 가장 가깝기 때문에 동일한 음운을 하나의 대표스펙트럼으로 나타낼 수 있음으로 해서 주파수변동을 흡수했기 때문이라 생각된다.

그러나, 신경회로망에 학습된 집단의 수가 16인 경우, 인식율이 떨어지는 것은 집단의 수가 음운의 수보다 너무 많기 때문에 주파수 변동으로 인해 발생하는 변동된 스펙트럼까지 패턴에 포함되어 유사도를 측정할 때에 오인식되는 요인이 된 것으로 생각되며,  $\alpha$ 값이 3인데도 인식율이 급격히 떨어지지 않는 것은 신경회로망에 의해 주파수 변동의 폭을 흡수한 후 멤버쉽함수를 작성했기 때문인 것으로 생각된다.

표 5. 뉴럴-퍼지패턴매칭에 의한 인식율

Table 5. Recognition rates by Neural-Fuzzy pattern matching.

[%]

$\alpha$ CLUSTERS SPEAKER	1			2			3		
	4	8	16	4	8	16	4	8	16
MA	100.0	100.0	100.0	100.0	100.0	100.0	98.2	100.0	98.2
MB	100.0	100.0	100.0	100.0	100.0	100.0	98.2	98.2	98.2
FC	100.0	98.2	96.4	100.0	98.2	96.4	98.2	98.2	94.6
AVG.	100.0	99.4	98.8	100.0	99.4	98.8	98.2	98.8	97.0

표 6. 인식방법에 따른 기억용량, 계산량 및 인식율의 비교

Table 6. Comparison of memory capacity, computational requirements, recognition rates for recognition methods.

METHODS	MEMORY CAPACITY	COMPUTATIONAL REQUIREMENTS	RECOGNITION RATES [%]
DTW	$B \times P \times K \times \sum_{i=1}^{L_R} F_i$	$F \times N_w \times L \times (P^2 + P) \times 2 \times K$	97.6
F.P.M	$1 \times P_B \times \sum_{i=1}^{L_R} F_i$	$F \times L \times P_B \times 2$	94.1
N-F.P.M.	$1 \times P_B \times \sum_{i=1}^{L_R} F_i$	$F \times L \times P_B \times 2$	98.8

3. 인식성능의 검토

음성인식에서 기억용량과 계산량 및 인식율과의 관계는 음성인식의 효율성을 평가하는 주요 요소이다. 본 절에서는 뉴럴-퍼지 패턴매칭의 기억용량, 계산량, 및 인식율을 종래의 DTW 방법과 비교하여 뉴럴-퍼지 패턴매칭에 의한 음성인식 방법을 검토 고찰하고자 한다. 표 6에는 3가지 음성인식 방법에 따른 기억용량, 계산량 및 본 연구의 음성데이터에 대한 인식율을 나타내었다. 이표에서 F.P.M.은 퍼지패턴매칭이며, N-F.P.M.은 본연구의 뉴럴-퍼지패턴매칭이며 인식율은 표 2, 표 4 및 표 5를 참조하였다.

또한, B는 특징파라미터 요소를 나타내는 비트수, P 차수, P<sub>B</sub>는 뉴럴-퍼지 패턴매칭의 차수, L<sub>B</sub>는 표준패턴으로 등록된 단어의 수, F<sub>i</sub>는 i번째 단어의 프레임 수, N<sub>w</sub>는 DTW의 창 길이, K는 멀티 템플릿의 수이다. 본 연구의 실험에서는 P<sub>B</sub>=18, B=32, P=14, K=3, N=15으로 하였다. 이 때의 기억용량의 비는 약 1/50 계산량의

비는 약 1/700 이 되어 뉴럴-퍼지 패턴매칭에서 필요한 계산량이 작음을 알 수 있다.

이에 반해 같은 실험 조건과 같은 음성데이터를 대상으로 뉴럴-퍼지패턴매칭에 의한오인식은 DTW방법 및 퍼지패턴매칭방법의 오인식보다 감소되었다. 따라서 뉴럴-퍼지패턴매칭에 의한 음성인식에서는 작은 기억용량과 계산량으로 종래의 DTW방법 및 퍼지패턴매칭에 의한 음성인식보다 높은 인식율을 얻으므로써 그 성능의 우수성을 확인하였다.

VI. 결 론

본 연구에서는 음성의 주파수변동과 시간변동 문제를 보완하기 위하여 신경회로망과 퍼지이론을 결합한 뉴럴-퍼지 패턴매칭에 의해 특정화자 고립단어인식을 수행하였다. 이 방법을 평가하기 위하여 28개의 도시명을 대상으로 음성인식 실험을 수행한 결과 다음과 같은 결론을 얻었다.

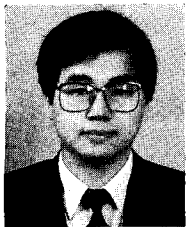
- (1) 뉴럴-퍼지패턴매칭의 음성인식에서 같은 기억용량과 계산량으로 퍼지패턴매칭의 것보다 오인식을 감소시키므로써 신경회로망에 의한 패턴작성법의 유효성을 확인하였다.
- (2) 뉴럴-퍼지패턴매칭에서 종래의 DTW 방법에 의한 음성인식방법보다 작은 기억용량과 계산량으로 높은 인식율을 얻으므로써 인식성능의 우수성을 확인하였다.

參 考 文 獻

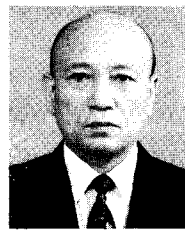
- [1] W. A. Lea, *Tends in speech recognition*, Prentice-Hall, Inc. 1980
- [2] H. Sakoe, S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-26, Feb. 1978
- [3] S. E. Levinson, L. R. Rabiner, A. E. Rosenberg, J. G. Wilpson, "Interactive clustering techniques for selecting independent reference templates for isolated word recognition," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-27, Apr. 1979
- [4] H. Sakoe, R. Isotani, K. Yoshida, "Speaker-independent word recognition using dynamic programming neural network," *ICASSP 89*, S1. 8, 1989
- [5] A. Waibel, T. Hanazawa, G. Hinton, K.

- Shikano, K. Lang, "Phoneme recognition using Time-Delay Neural Networks," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-37, 1989
- [6] 崔元奎, 秋月影雄, "ファシイ推論による母音認識と韓國語連續音聲への應用," *電學論 C*, 111-5, 1991
- [7] J. Fuzimoto, T. Nakatani, M. Yoneyama, "Speaker-independent word recognition using fuzzy pattern matching," *Fuzzy Sets and Systems* 32, North-Holland, 1989
- [8] S. Nakanishi, T. Takagi, "Pattern recognition by neural network and fuzzy inference," *Proc. ICFNN*, 1990
- [9] 崔元奎, 李羲, 秋月影雄, "韓國語音聲認識のための改良形ホルマントトラッキングと 數子音聲認識への應用," *電學論 C*, vol. 109-C, no. 10, 1988
- [10] 松本博, 中川正雄, 米山正秀, "ローカルヒーク荷重平均辭書を用いた不特定話者單語 音聲認識," *電子通信學會論文誌*, vol. J68-A, no. 1, 1985
- [11] R. A. Cole, *Perceptron and production of fluent speech*, Lawrence Erlbaum Associates, Inc., 1980.
- [12] Y. Linde, A. Buzo, R. M. Gray, "An algorithm of vector quantization design," *IEEE Trans. Communication*, vol. COM-28, 1980.

著 者 紹 介



**李 基 榮 (正會員)**  
 1961年 5月 7日生. 1984年 2月 명지대학교 전자공학과 졸업. 1986年 2月 명지대학교 전자공학과 공학석사. 1992年 2月 명지대학교 전자공학과 공학박사



**崔 甲 石 (正會員)**  
 1930年 9月 30日生. 1955年 8月 서울대학교 공과대학 통신공학과 졸업. 1975年 2月 25日 명지대학교 공학박사. 1980年 8月~1981年 8月 일본동경대학교 공학부 전자공학과 연수. 현재 명지대학교 정보통신공학

과 교수



# Correction to “Design of Adaptive Regulator Using the Explicit Criterion Minimization”

Sang Jae Lee

In the above paper<sup>1</sup>, the following corrections should be incorporated. For Eqs. (3), (5), (10), (12), (13) and (14) and Fig. 1 and 2, reference[1] must be quoted.

I deeply regret to Prof. Sternad that I did not quote his earlier work[1] on adaptive control using explicit criterion minimization, in the references of the above paper<sup>1</sup>. The major difference between the two papers is that a major part of the above paper<sup>1</sup> is devoted to establishing a useful algorithm for criterion minimization using a recursive prediction error method, proposing a choice of input penalty, and making a simulation using an independent method even though the above paper<sup>1</sup> used the same examples as reference[1]. I had also obtained different regulator performances. Nevertheless, reference[1] is an original and valuable contribution to the subject and should have been cited.

## ACKNOWLEDGEMENT

I am grateful to Prof. Sternad of Uppsala University for his valuable paper[1] which was, regrettably, missed in the references of my above paper<sup>1</sup>. And I really apologize to him that I have used his earlier and valuable work without his permission. I recognize that such a thing is a violation of research ethics. So I promise that I will never do such a thing hereafter, with an apology to Prof. Sternad and Korean Institute of Telematics and Electronics.

## REFERENCE

1. M. Sternad, “Optimal and Adaptive Feedforward Regulators”, Ph. D. Thesis, Uppsala University, Sweden, 1987.

---

The author is with the Dept. of Electronics & Control Engineering, Kum-Oh National Institute of Technology, Kyungbuk, Korea.

<sup>1</sup> Sang Jae Lee, *et al.*, *Journal of Korean Inst. of Telematics and Electronics*, vol. 27, no. 7, pp. 16-23, 1990.

# ON RESEARCH ETHICS AND THE COPYING OF SCIENTIFIC RESULTS

Mikael Sternad\*

May 1992, revised September 1992

## Abstract

This short note is a comment on the following subject: what are the rules for copying results from the published and unpublished papers of other researchers? When is such a practice allowed, and when does it constitute a violation of research ethics, a case of plagiarism?

Submitted to the Journal of the Korean Institute of Telemechanics and Electronics.

This note has been written because there is a definite need of discussing some rules of scientific conduct. The more these rules become known, the less is the risk for mistakes, embarrassment and mistrust within the scientific community.

My interest in these matters is partly a result of a mistake, made in a recent paper[4]. In that paper, the explicit criterion minimization method of adaptive control, suggested by Trulsson and Ljung[5], is generalized to a feedback-feedforward controller structure. A part of that paper closely follows my own work [1]-[3]. Surprisingly, no references whatsoever were given to [1]-[3].

In a subsequent correspondence with the authors, the following things became clear.

- The first author of [4] admitted that he had made an error in not quoting my results, which he had in his hands as he was writing his paper. He immediately apologized sincerely for it, and I accept his apologies. At the time of writing the paper, he seems not to have been quite aware that he was doing something improper
- The first author was solely responsible for the content of the paper. The names of the other three authors were added purely out courtesy: They had supplied some basic material for the paper. The second, third and fourth authors should thus not be held responsible for any errors.

---

\* Systems and Control Group, Department of Technology, Uppsala University, P O Box 27, S-751 03 Uppsala, Sweden.

## References

- [1] Sternad M  
"Optimal and Adaptive Feedforward Regulators", Ph. D. Thesis, Department of Technology, Uppsala University, Sweden, 1987.
- [2] Sternad M  
"Adaptive feedforward and disturbance decoupling based on explicit criterion minimization". Report UPTEC 8615R, Department of Technology, Uppsala University, Sweden, 1986.
- [3] Sternad M  
"Disturbance decoupling adaptive control". Proc 2nd IFAC Workshop on Adaptive Systems in Control and Signal Processing, Lund, Sweden, July 1986, pp. 399-404.
- [4] Sang Jan Lee, Chang Hyun Chae, Tae Chon Ahn and Si Hyung Cho, "Design of adaptive regulator using the explicit criterion minimization". Journal of the Korean Institute of Telemechanics and Electronics, vol. 27 (7), pp. 16-23, 1990.
- [5] Trulsson E and L Ljung  
"Adaptive control based on explicit criterion minimization". Automatica, vol. 21, pp. 385-399.