

# 다구간 등분할법과 퍼지추론을 이용한 단어인식

## Word Recognition Using Multi-section Equi-segmentation and Fuzzy Inference

최 승 호\*, 최 갑 석\*\*  
(Seung Ho Choi\*, Kap Seok Choi\*\*)

### 요 약

본 논문은 다구간 등분할법과 퍼지추론으로 단어인식을 행하는 패턴매칭법을 제안한다. 패턴매칭시 발생하는 시간변동은 발생순서에 따라 등간격으로 다구간 분할함으로써 해결하고, 주파수변동은 구간의 차수별로 정해진 퍼지관계로부터 패턴간의 퍼지추론이 행해짐으로써 흡수한다. 추론에 사용된 삼각형 멤버십함수의 중심값과 변동폭은 패턴의 평균값과 분산값으로 대응되도록 작성한다.

20대 남성 2인이 발성한 데이터를 사용하여, 제안된 방법으로 DDD지역명 28개를 구간수와 변동폭을 달리하여 인식실험한 결과, 8구간과 4배의 변동폭을 가질 때 92%의 인식율을 얻었다.

### ABSTRACT

In this paper, we propose a new pattern matching method which achieves word recognition by the multi-section equi-segmentation method and the fuzzy inference.

On pattern matching, the problem of time variation is solved by equally segmenting a pattern in the order of utterance and then frequency variation is solved by fuzzy inferring patterns from relationships between each LPC order of sections.

The mean and the variance of the pattern correspond to the center value and the variation width of the triangular membership function used in the fuzzy inference.

The proposed algorithm has been applied to the recognition for twenty eight DDD area names uttered by two men in the twenties in different numbers of sections and the variation widths.

In the eight sections and the four times variation width, the new approach showed the best result of 92% recognition.

### I. 서 론

인간상호간의 기본적인 의사전달 수단인 음성은 정보통신 기술의 발전에 따라 인간과 기계간의 정보

교환 수단으로 이용되고, 사용언어가 각 나라마다 다르기 때문에 자국어에 대한 독자적인 연구가 필요하다[1].

주로 패턴매칭방법이 사용되는 음성인식시스템에서는 음성의 음향적 특성이 잘 나타나는 특징량으로 패턴을 작성하지만, 패턴화하는 과정에서 시간변동과 주파수변동 문제가 발생된다. 전자는 화자에 따른 발성속도와 동일화자가 동일단어를 발성했을 때 발성지속시간차에 의한 변동이며, 후자는 발성상황과

\*Dept. of Information & Communication Eng. Dong Shin Univ.

동신대학교 정보통신공학과

\*\*Dept. of Information & Communication Eng. Myung Ji Univ.

명지대학교 정보통신공학과

접수일시: 1993년 5월 13일

화자마다 성도의 깊이가 달라짐에 따라 발생하는 변동이다[2].

이러한 변동들을 최소화 하기 위해서 Fujimoto는 Zadeh가 제안한 퍼지집합의 개념을 이용하여 시간변동과 주파수변동을 동시에 해결하는 인식장치를 실용화하였다[3]. 그는 성도의 주파수특성을 고려해서 피크의 존재 유무를 1과 0으로 표현한 BTSP(binary time spectrum pattern)을 작성하여 시간변동은 선형팽창법을 사용하고, 주파수변동은 각 패턴별로 멤버쉽함수를 작성하여 그가 정의한 유사도를 이용해서 해결하였다.

최근의 비지추론에 의한 단어인식에서는 각 음소를 포먼트, 피치, 에너지 등의 특징파라미터들로 나타내어 개인차에 의한 각 특징파라미터들의 변동량을 퍼지수로 놓고, 이로부터 멤버쉽함수를 작성하여 그 변동을 표현한 뒤 각 특징파라미터들과 각 음소간의 퍼지관계합성과 퍼지규칙으로 지식법에 의한 퍼지추론을 행하여 주파수변동 문제를 해결하였다[4].

이와같이 퍼지개념을 이용한 단어인식에서는 변동폭을 줄이기 위해 많은 시간을 예비실험에 소요할 뿐만 아니라 각 특징파라미터에 고정된 변동폭을 주어 모든 데이터를 처리하기 때문에 패턴변화에 능동적으로 대처하지 못하는 단점이 있다[3].

본 연구에서는 이러한 점에 착안하여, 음성의 음향적 특성이 유지되도록 패턴을 발생순서에 따라 등간격으로 다구간 분할하고, 다구간 등분할된 표준패턴과 미시패턴으로 각 단어의 구간별 차수마다 퍼지관계를 정하여 각 구간의 퍼지관계합성과 퍼지규칙에 의해서 패턴간의 퍼지추론을 행하는 방법을 제안하여 패턴변화에 능동적으로 대처하고, 구간수와 삼각형 멤버쉽함수의 변동폭을 달리하면서 패턴간의 오인식율을 고찰한다.

### II. 다구간등분할법

인간의 청각적 특성에서, 모음의 평균발생주기는 대부분 70ms 전후로 안정되어 있으나, 자음의 경우는 음소에 따라 5~130ms정도의 변화폭을 갖음으로 음절 주기를 크게 변화 시키는 원인이 된다. 또한, 단어들은 대부분 한 음절당 3~4개의 음소로 구성되어 있다[5]. 이러한 점에 착안하여 본 연구에서는 패턴의 시간길이를 정규화시키고, 개인차에 따른 시간변동을 효과적으로 처리하기 위해 그림 1과 같은 절차로 다구간 등분할법을 제안하여 사용한다.

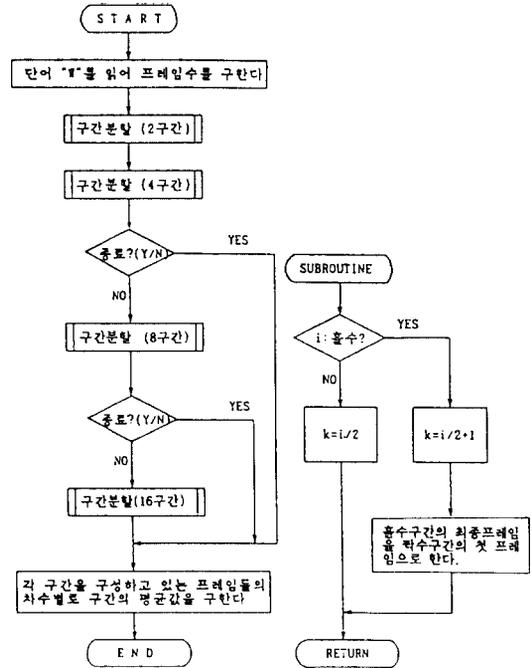
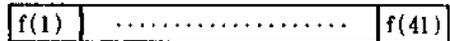


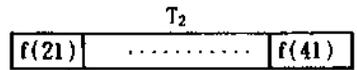
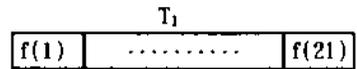
그림 1. 다구간 등분할법의 작성 흐름도.  
Fig 1. Flowchart of the multi-section equi-segmentation method.

예를 들어, 음성길이 41프레임인 하나의 음성패턴에 대해 등간격으로 여덟 구간 나누는 과정은 다음과 같다.

#### 1. 음성의 시간길이가 41프레임일때

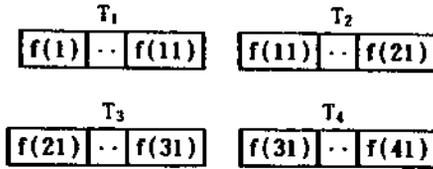


2. 첫번째 구간의 21번째 프레임을 두번째 구간의 첫 프레임에 삽입하여 두구간으로 나눈다. 두번째 구간에 21번째 프레임을 중첩시키므로써 두구간의 프레임수를 동일하게 한다.

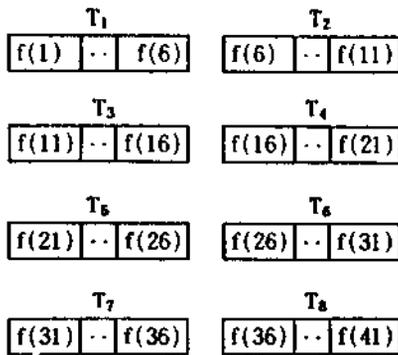


3. 첫번째 구간의 11번째 프레임을 두번째 구간의 첫 프레임에 삽입하고, 세번째 구간의 31번째 프레임을 네번째 구간의 첫 프레임에 삽입한다. 11번째와 31번째

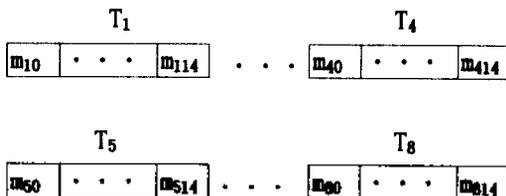
프레임을 중첩시키므로써 네구간의 프레임수를 같게 한다.



4. 첫번째 구간의 6번 프레임을 두번째 구간의 첫 프레임에, 세번째 구간의 16번 프레임을 네번째 구간의 첫 프레임에, 다섯번째 구간의 26번 프레임을 여섯번째 구간의 첫 프레임에, 일곱번째 구간의 36번 프레임을 여덟번째 구간의 첫 프레임에 각각 삽입한다. 6, 11, 16, 26, 36번째 프레임을 중첩시키므로써 여덟구간에 프레임수를 동일하게 한다.



5. 1~4과정을 통해 여덟구간으로 분할되면서 6, 11, 16, 21, 26, 31, 36번 프레임을 삽입하기 때문에 각 구간은 6개의 프레임으로 각각 구성되어 등간격으로 다구간 분할된다. 이때, 각 프레임들은 15개의 LPC계수를 갖기 때문에 각 구간은 90개의 계수를 갖는다.
6. 90개의 계수들로 구성된 구간들은 각 프레임의 같은 차수끼리 계수들을 합하여 구간별로 15개의 평균값을 각각 갖는다. 이러한 평균값은 그 구간을 대표하는 차수별 대표값이 된다.



결국, 각 구간은 LPC 계수의 차수별 평균값은 다음과 같이 구해진다.

$$m_{JK} = \frac{1}{6} \sum_{i=1}^6 A_{KJ}, \quad J=1, 2, \dots, 8, \quad K=0, 1, \dots, 14 \quad (2-1)$$

여기에서 J는 구간수, K는 LPC계수의 차수 그리고 f는 프레임을 나타낸다.

따라서, 1~4과정으로 부터 작성된 패턴들은 규칙적인 방법으로 프레임을 삽입하기 때문에 정보손실을 방지하고, 5~6과정에서는 각 구간의 차수별로 평균값을 구하기 때문에 정보압축이 이루어진다[6].

### III. 표준패턴 작성

표준패턴은 다구간 등분할법에 의해 작성된 패턴들을 선형결합하여 구간의 차수별 평균값과 분산값으로 구성되며, 단어당 하나의 표준패턴을 갖는다. 단어 W에 대한 l번째 패턴의 J번째 구간의 차수별 평균값을  $m^l_{JK}$ 라 하면 단어 W에 대한 전체 평균값  $M^W_{JK}$ 와 분산값  $V^W_{JK}$ 는 식 (3-1)과 (3-2)와 같다.

$$M^W_{JK} = \sum_{l=1}^n \sum_{k=1}^n \left( \frac{1}{n} \sum_{i=1}^n m^l_{JK} \right), \quad l = 1, 2, \dots, 6 \quad (3-1)$$

$$V^W_{JK} = \sum_{l=1}^n \sum_{k=1}^n \left[ \frac{1}{n-1} \sum_{i=1}^n ((M^W_{JK} - m^l_{JK})^2)^{1/2} \right] \quad (3-2)$$

여기에서 l는 단어당 패턴수이다.

그림 2는 단어 “동두천”에 대한 표준패턴 작성과정의 한 예이다. “동두천”의 발성인 2인의 화자가 3회 발성한 6개로 각 발성길이는 87, 44, 79, 68, 49, 66 프레임이고, 이때, 단어를 여덟구간으로 등분할 하면 발성길이에 따라 프레임의 삽입수가 각각 다르다. 즉, 87프레임은 1프레임을, 44프레임은 4프레임을, 79프레임은 1프레임을, 68프레임은 4프레임을, 49프레임은 7프레임을, 66프레임은 6프레임을 각각 여덟구간으로 등분할되면서 프레임을 삽입하게 된다. 각 패턴들은 여덟구간으로 등분할되어 음성의 시간길이에 무관하게 일정한 패턴길이가 된다. 즉, 각 단어의 표준패턴은 각각 120개의 평균값과 분산값으로 작성된다.

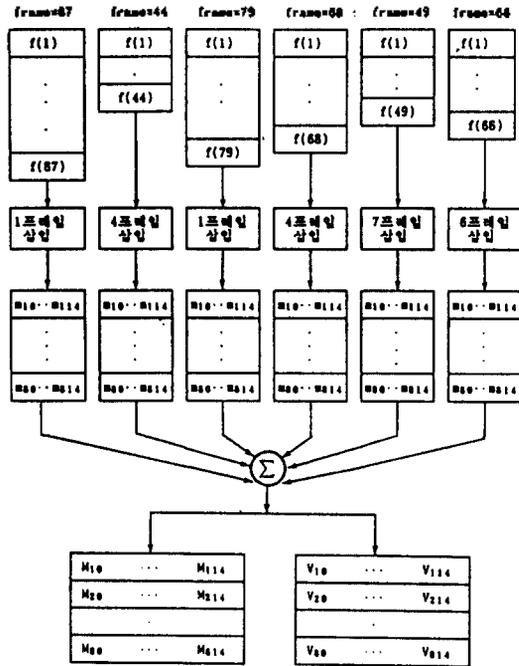


그림 2 다구간 등분할법으로 작성된 단어 "동두천"의 표준패턴.  
Fig 2. Reference pattern of "Dongducheon" by the method of multi-section equi-segmentation.

IV. 음성의 멤버쉽함수

선형결합된 표준패턴은 패턴의 공통적특성을 갖는 평균값과 패턴의 변화량을 나타내는 분산값으로 구성되는 점에 착안하여 음성패턴에 삼각형 멤버쉽함수를 사용한다. 이때, 패턴의 평균값과 분산값을 멤버쉽함수의 중심값과 변동폭으로 각각 대응시킨다.

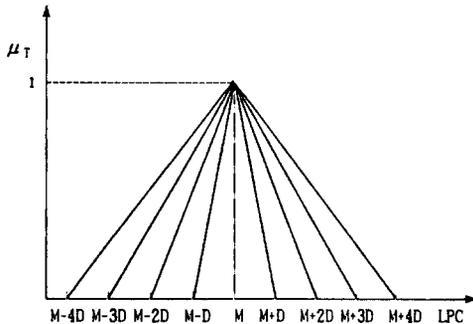


그림 3. 변동폭에 따라 삼각형 멤버쉽함수 (평균값 = M, 분산값 = D).  
Fig 3. Triangle membership function for the variation width (mean = M, variance = D).

또한, 화자의 변동차를 최소화하기 위해 변동폭의 범위를 4 배까지 조절한다. 그림 3은 규칙적으로 변동폭을 4 배까지 조절된 범위를 나타낸 것이다.

표 1은 표준패턴 "서울"의 두번째 구간을 나타낸 예이다. 이를 고찰해보면 차수가 낮을수록 분산값은 커지며, 차수가 높을수록 분산값은 작아지고 있다. 즉, 차수가 낮은 부분은 거의가 저주파성분이고 변동폭이 넓어서 화자의 변동에 기여할 수 있는 정도가 높고, 차수가 높은 부분은 거의가 고주파성분이고 변동폭이 좁아서 화자의 변동에 기여할 수 있는 정도가 작다는 것을 의미한다. 이 결과로부터 차수에 따라 변동폭이 달라지기 때문에 멤버쉽함수의 변동폭은 구간의 차수별로 다르다.

표 1. 표준패턴 "서울"의 2번째 구간의 평균값과 분산값.  
Table 1. Mean and variance values in the 2nd section of the reference pattern "Seoul".

차 수	평 균 값	분 산 값
A <sub>0</sub>	1.00000	0.00000
A <sub>1</sub>	0.76027	0.64738
A <sub>2</sub>	0.10820	0.41561
A <sub>3</sub>	- 0.25355	0.12639
A <sub>4</sub>	- 0.20708	0.18883
A <sub>5</sub>	- 0.04260	0.05034
A <sub>6</sub>	0.12020	0.02657
A <sub>7</sub>	- 0.20300	0.05414
A <sub>8</sub>	- 0.08186	0.02579
A <sub>9</sub>	0.02319	0.01101
A <sub>10</sub>	- 0.06216	0.01295
A <sub>11</sub>	0.13836	0.02492
A <sub>12</sub>	- 0.04795	0.02524
A <sub>13</sub>	- 0.01898	0.01150
A <sub>14</sub>	- 0.06483	0.00193

지금까지의 퍼지이론을 이용한 음성인식에서는 동일한 특징파라미터에 같은 크기의 변동폭을 주기 때문에 패턴변화에 능동적이지 못하여 오류가 발생할 수 있다[7]. 이러한 문제점을 극복하기 위해, 앞서 고찰한 결과로 부터 각 구간의 차수에 따라 멤버쉽함수의 변동폭을 동적으로 부여함으로써 화자에 따른 변동차를 효과적으로 좁을 수 있다.

V. 퍼지추론

표준패턴과 미지패턴의 멤버쉽함수들이 패턴간의

퍼지관계와 합성에 의해 다중 퍼지추론을 행하는 패턴매칭법을 제안하여 단어인식을 한다. 퍼지추론으로 패턴매칭을 행하는 과정은 다음과 같다. 표준패턴의 단어당 동분할된 J번째 구간  $T_j$ 에서 K번째 차수의 평균값인  $M_{JK}^w$ 을  $P_K$ 라 하면  $P_K$ 의 집합을 X라 하고, X의 퍼지부분 집합을 A라 할때 A와 X사이에는 식 (5-1)과 같은 관계가 성립한다.

$$A \subset X \tag{5-1}$$

미지패턴의 각 단어당 구간  $T_j$ 의 집합을 Y라 하면, X와 Y상의 퍼지관계는 식 (5-2)와 같이 정의된다.

$$R \subset X \times Y \tag{5-2}$$

이로부터,  $X \times Y$ 의 퍼지관계 R, X와 그 부분집합 A의 퍼지관계를 S라 하면, R과 S의 합성 ROS는 멤버쉽함수로 구해지는 퍼지관계이며, 식 (5-3)과 같다.

$$ROS \Leftrightarrow \max_{\mu_{ROS(T_j)} P_K} [\mu_R(P_K, T_j) \wedge \mu_A(P_K)] \tag{5-3}$$

여기서  $\mu_R$ 과  $\mu_A$ 는 각각 퍼지관계 R과 부분집합 A의 멤버쉽함수이다.

따라서, 위에서 서술한 퍼지관계와 합성의 예는 다음과 같다. 미지패턴 구간의 K번째 차수와 어떤 단어의 첫번째 구간의 K번째 차수의 확신도는 퍼지관계의 합성 ROS에 의해 max-min 연산으로 구해지며 그 과정은 그림 4와 같으며, 표준패턴은 평균값이 0.7, 분산값이 0.1이고, 미지패턴은 평균값이 0.75일때

의 확신도 산출과정이다. 여기서 미지패턴의 분산값은 표준패턴의 분산값인 0.1을 사용한다.

각 단어의 구간에서 특징파라미터의 변동과 전건부의 퍼지변수는 구간의 차수별로 퍼지규칙을 적용한다. 다중 퍼지추론에서 후건부는 구간의 확신도이고, 전건부는 평균연산(&)으로 결합된 특징파라미터  $P_K$ 와 구간  $T_j$ 사이의 퍼지관계  $R_{JK}$ 이다.

따라서 본 연구에서 사용된 퍼지추론을 위한 퍼지규칙의 예는 다음과 같다.

“서술”

규칙 1.: if  $P_{10}$  is  $R_{10}$  &  $P_{11}$  is  $R_{11}$  ... &  $P_{1K}$  is  $R_{1K}$  Then  $T_1$

규칙 2.: if  $P_{20}$  is  $R_{20}$  &  $P_{21}$  is  $R_{21}$  ... &  $P_{2K}$  is  $R_{2K}$  Then  $T_2$

⋮ ⋮ ⋮

규칙 J.: if  $P_{J0}$  is  $R_{J0}$  &  $P_{J1}$  is  $R_{J1}$  ... &  $P_{JK}$  is  $R_{JK}$  Then  $T_j$

결국, 주어진 추론규칙에 의해 인식대상단어를 판정하기까지의 퍼지추론과정을 살펴보면 다음과 같다.

1) 각 구간의 차수별 확신도

구간  $T_j$ 에서 특징파라미터  $P_K$ 의 확신도는 퍼지관계합성에 의해 식 (5-4)와 같이 나타난다.

$$R_{JK} \circ S_K = \max_{P_K} [\mu_{R_{JK}}(P_K, T_j) \wedge \mu_{S_K}(P_K)] = W_{JK} \tag{5-4}$$

따라서,  $W_{JK}$ 는 J번째 구간의 K번째 차수의 계수값에 대한 확신도를 나타낸다.

2) 구간별 확신도

앞에서 구해진 각 구간의 차수별 확신도  $W_{JK}$ 를 이용하여 단어의 구간  $T_j$ 에 대한 구간별 확신도  $W_j$ 는 식 (5-5)와 같이 평균확신도로 정의된다.

$$W_j = \frac{1}{n+1} \sum_{k=0}^n W_{JK} \tag{5-5}$$

평균확신도는 구간  $T_j$ 에 대해 각 구간별 확신도를 산출하는 과정에서 한 개의 특징파라미터가 잘못 연산되어 구간  $T_j$ 의 확신도로 구해지는 것을 방지하기 위함이다[4].

이상과 같이 등간격으로 다구간 분할된 표준패턴과 미지패턴간의 멤버쉽함수들이 패턴매칭하는 과정을 행렬의 형태로 표현하면 그림 5와 같다.

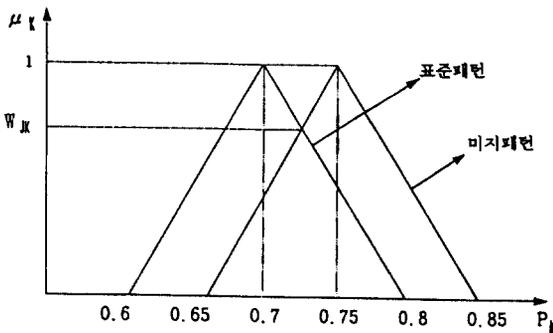


그림 4. 퍼지관계와 합성에 의한 확신도의 산출.  
Fig 4. Decision of certainty factor by composition of fuzzy relations.

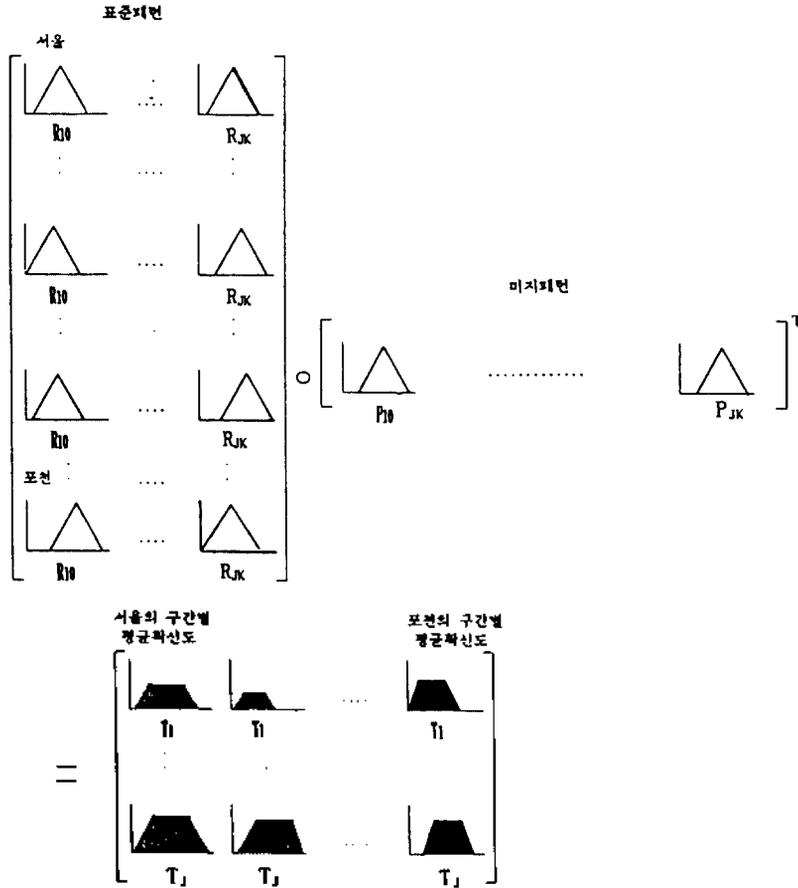


그림 5. 다구간으로 등분화된 표준패턴과 미지패턴간의 퍼지추론.

Fig 5. Fuzzy inference between reference pattern and unknown pattern in the multi-section equi-segmentation.

3) 단어별 최종확신도

구해진 평균 확신도  $W_j$ 를 이용하여 구해지는 단어에 대한 최종확신도  $H_w$ 는 식 (5-6)과 같다.

$$H_w = \frac{1}{n} \sum_{i=1}^n W_i \tag{5-6}$$

4) 인식단어의 판정

미지패턴과 표준패턴과의 최종확신도가 가장 큰 것을 인식단어  $Q_w$ 로 판정한다.

$$\max_{1 \leq w} \{H_w\} \Rightarrow Q_w \tag{5-7}$$

VI. 인식실험 및 고찰

1) 분석조건

본 연구에서 사용된 데이터는 DDD지역명 28개(직할시 이상 지역명 6개, 경기도 지역명 22개)를 20대 이상의 성인남성 2인이 각각 5회씩 발성한 280개이다. 이들중 표준패턴은 2인이 3회씩 발성한 데이터로 작성하고, 미지패턴은 표준패턴에서 사용되지 않은 나머지 2인이 2회씩 발성한 112개 데이터를 사용한 다.

음성패턴을 작성하기 위한 전처리 작업으로, 먼저 끝점을 검출하고 선형예측분석을 한다[8]. 그 분석조건은 표 2와 같다.

표 2. 분석조건.

Table 2. Analysis conditions for speech signals.

Sampling frequency	: 8	[kHz]
cut-off Freq. of LPF	: 3.4	[kHz]
A/D Resolution	: 12	[bit]
Analysis Window	: 20	[ms]
Overlapping interval	: 10	[ms]
Window function	: Hamming Window	
LPC order	: 14	

2) 실험방법

제안된 다구간 등분할법과 퍼지추론으로 단어인식 실험을 행하고, 이때 구간의 수(4, 8, 16)와 삼각형 멤버함수의 변동폭(1D, 2D, 3D, 4D)을 달리 하여 비교하였다. 다구간으로 등분할된 패턴들 사이에서 이루어지는 퍼지추론의 단어인식 구성도는 그림 6과 같다.

표 3은 같은 발성위치에 동일한 음절을 갖고 있는 미지패턴들을 비교한 것이며, 여덟번째 구간 중에서 여섯번째 구간의 차수별 평균값을 나타낸 예이다.

표 4와 표 5는 여덟번째 구간 중에서 여섯번째 구간에 대한 표준패턴의 차수별 평균값과 분산값을 인천, 이천, 동두천, 포천등과 비교하여 나타낸 예이다.

또한, 그림 7은 패턴변화를 인천, 이천, 동두천, 포천등과 비교하기 위하여 구간의 차수별 평균값으로

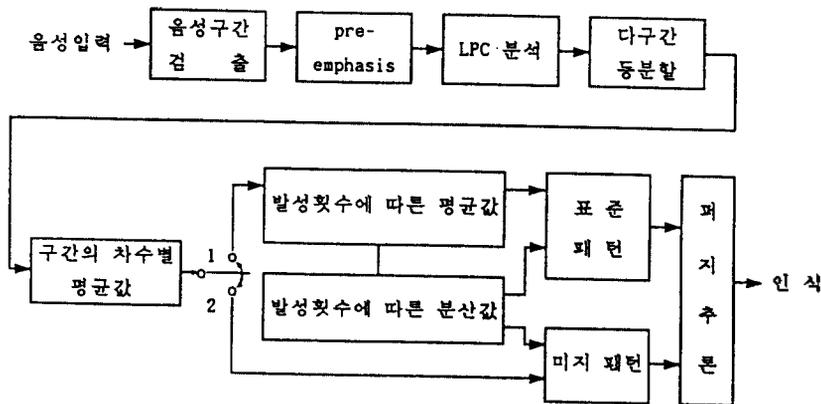
구한 스펙트럼포락선을 첫번째 구간에서 여덟번째 구간까지 나타낸 것이다. 이로부터 같은 발성위치에 동일음절이 있는 패턴들은 거의 유사한 포락선을 보이고 있음을 알 수 있었다.

퍼지합성의 연산규칙에 의해 구간의 차수별 확신도가 구해지고, 구해진 차수별 확신도를 평균하여 구간의 확신도가 정해진다. 이때 같은 발성위치에 동일

표 3. 미지패턴 평균값의 예(6번째 구간).

Table 3. An example of mean of test pattern(6th section).

패턴 차수	인 천	이 천	동두천	포 천
0	1.000000	1.000000	1.000000	1.000000
1	-1.12897	-0.84832	-0.72566	-1.07583
2	-0.51538	-0.48708	-0.71569	-0.61090
3	0.588861	0.306758	0.059325	0.483769
4	0.477118	-0.18353	0.714928	0.774624
5	0.196190	0.070510	0.388570	0.260330
6	-0.79569	0.192214	-0.61027	-0.74424
7	-0.15435	0.137256	-0.36616	-0.50876
8	0.265131	-0.12969	-0.11382	0.187001
9	0.344646	0.083846	0.469969	0.601796
10	-0.02216	-0.04218	0.317418	0.112904
11	-0.19217	-0.10728	-0.16884	-0.24809
12	-0.00156	0.045542	-0.29651	-0.23007
13	0.008614	0.044407	0.012392	0.051304
14	0.008429	0.008987	0.143966	0.080601



1: Generation reference pattern  
2: Recognition

그림 6. 다구간 등분할법과 퍼지추론에 의한 인식구성도.

Fig 6. Block diagram of recognition using the method of multi-section equi-segmentation and fuzzy inference.

표 4. 표준패턴의 평균값.

Table 4. Mean of pattern.

패턴 차수	인 천	이 천	동두 천	포 천
0	1.000000	1.000000	1.000000	1.000000
1	-0.94294	-0.81466	-0.58707	-1.89687
2	-0.47078	-0.31171	-0.47501	-0.49090
3	0.030584	0.098668	-0.15044	0.215431
4	0.827999	0.22640	0.574710	0.53323
5	0.240085	0.104087	0.308816	0.209386
6	-0.62288	0.109659	-0.38941	0.48897
7	-0.37552	0.139400	-0.32930	-0.22830
8	0.053274	0.004754	-0.14775	0.015733
9	0.646439	0.059506	0.448694	0.414865
10	-0.01109	-0.09979	0.298593	0.064855
11	-0.27128	-0.08944	-0.17131	-0.20230
12	-0.14228	0.046292	-0.28561	-0.11114
13	0.105202	0.094751	-0.01293	0.030327
14	0.037979	-0.02992	0.140421	0.067113

표 5. 표준패턴의 분산값.

Table 5. Variance of reference pattern.

패턴 차수	인 천	이 천	동두 천	포 천
0	0.000000	0.000000	0.000000	0.000000
1	0.122447	0.097885	0.656454	0.189259
2	0.220454	0.228833	0.209402	0.102965
3	0.115179	0.279425	0.234470	0.289831
4	0.308248	0.188310	0.470436	0.168405
5	0.150423	0.085665	0.256697	0.031121
6	0.234760	0.107584	0.246013	0.134798
7	0.216638	0.129967	0.284193	0.097841
8	0.115827	0.039816	0.101454	0.200021
9	0.161122	0.085560	0.182826	0.081242
10	0.145716	0.068978	0.174202	0.143407
11	0.084111	0.077132	0.096609	0.049957
12	0.154926	0.058028	0.088606	0.228608
13	0.085775	0.018212	0.080322	0.101557
14	0.057595	0.036378	0.059998	0.083211

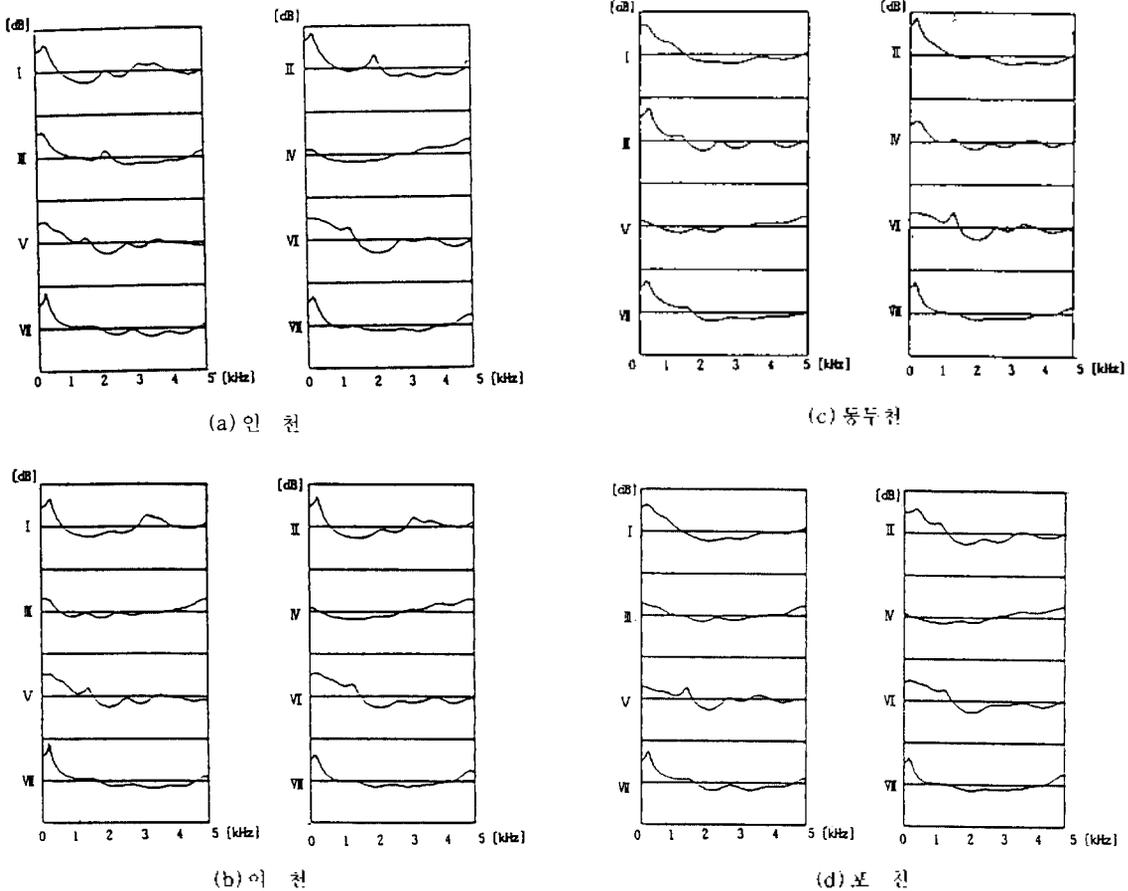


그림 7. 표준패턴들의 스펙트럼포락선.  
Fig. 7. Spectrum envelope of reference patterns.

표 6. 구간별 확실도의 예.

Table 6. An example of certainty factor in each section(6th section).

패턴 차수	인 천	이 천	동 두 천	포 천
0	0.00000	0.00000	0.00000	0.00000
1	0.66759	0.71234	0.89506	0.58096
2	0.91761	0.57162	0.66763	0.37473
3	0.97616	0.90500	0.94059	0.92428
4	0.69316	0.46494	0.83231	0.95278
5	0.94885	0.85683	0.98251	0.77878
6	0.73279	0.17674	0.84965	0.67494
7	0.80572	0.76884	0.62366	0.46845
8	0.85560	0.94130	0.90815	0.83822
9	0.43107	0.57122	0.80870	0.88000
10	0.97438	0.55171	0.76519	0.21211
11	0.15991	0.68571	0.76262	0.82436
12	0.95235	0.84686	0.71414	0.43095
13	0.66064	0.61159	0.77220	0.79667
14	0.82345	0.73834	0.74885	0.00000

음절이 있는 패턴에서 구한 여섯번째 구간의 차수별 확실도를 표 6에 나타내었다.

제한된 방법으로 구간수를 4, 8, 16개로, 삼각형 멤버함수의 변동폭을 1D, 2D, 3D, 4D로 할 때 그 인식률을 표 7에 나타내었다.

표 7. 다구간수와 변동폭에 따른 인식률.

Table 7. Recognition rate with multi-sections and variable widths.

변동폭	구간		
	4구간	8구간	16구간
1 D	60.7	64.3	74.1
2 D	83.9	87.5	82.1
3 D	85.7	89.3	85.7
4 D	87.5	92.0	87.5

표 7에서 보는 바와 같이 구간수가 8개인 경우가 4, 16개인 경우보다 인식률이 높은 것으로 나타났다. 즉, 등간격으로 다구간 분할될 때 구간수는 단어가 2음절로 이루어진 경우에 8개가 최적임을 알 수 있었다. 이는 실험에 사용된 단어가 6~8개의 음소로 구성되어 있기 때문에 8개의 구간으로 분할되는 것이 효과적이며, 또한, 퍼지추론에 사용된 삼각형 멤버함수의 변동폭은 분산값의 4배일때 가장 좋은 인식율이 나타난 것은 가우시안분포에서 분산값이 4배일때 모든 사건이 일어날 경우를 포함하기 때문이다.

오인식은 음향적 특성이 서로 다를 때 발생하는 경

우와 같은 발생위치에 동일한 음절이 있을 때로 분류할 수 있으며, 이러한 오인식들을 분석하기 위해 여덟 구간으로 등분할된 단어들의 변동폭을 1D에서 4D까지 고찰하여 표 8에 나타내었다. 표에서 보면, 전자는 패턴인식에서 주로 나타나는 현상으로써 변동폭이 4D일때 오인식율이 최소화 되었고, 후자는 패턴을 다 구간 등분할 때 나타나는 오인식으로서 변동폭이 4D일 때 가장 효과적임을 알 수 있었다.

표 8. 다구간 등분할법과 퍼지추론에 의한 오인식결과.

Table 8. False recognition result of the multi-section equi-segmentation method and fuzzy inference.

[오인식수/전체의 미지패턴 수]

오인식경우	변동폭 1D	2D	3D	4D
A	38/112	14/112	12/112	9/112
B	12/112	6/112	6/112	4/112
C	26/112	8/112	6/112	5/112

A : 오인식이 발생하는 모든 경우

B : 같은 발생 위치에 동일음절이 있는 경우

C : 음향적 특성이 서로 다른 경우.

## Ⅶ. 결 론

본 연구는 다구간 등분할법과 퍼지추론으로 단어인식을 행하는 패턴매칭법을 제안하여 DDD지역명 28개에 대해서 인식실험을 하였다.

음성인식의 문제점인 시각변동과 주파수변동의 해결을 위해 전자는 발생순서에 따라 등간격으로 분할하여 패턴의 시간길이를 정규화 하고, 후자는 각 구간별 차수마다 주어진 퍼지관계로부터 패턴간의 퍼지추론이 행해졌다. 또한, 멤버함수의 변동폭은 표준패턴의 분산값을 이용함으로써 멤버함수가 미지패턴의 변화에 능동적으로 대처할 수 있었으며, 특히 시행착오를 거친 사전실험을 통해 멤버함수의 변동폭을 결정할 필요가 없는 장점이 있다. 그리고 여러개의 인식파라미터를 사용하지 않고, 선형예측계수만으로 이루어진 인식파라미터를 사용하여 간단한 음성인식 시스템을 구성할 수 있으며, 92%의 높은 인식률을 보임으로써 제안된 방법이 타당함을 알 수 있었다.

본 연구에서 제안된 방법은 대어휘 음성인식시스템 분야에서 적용될 것으로 사료되며, 앞으로 구간의 동적인 분할방법과 퍼지추론에 의한 효율적인 인식 과정에 대한 연구가 진행될 것이다.

## 참 고 문 헌

1. G. Bristow. "Electronic Speech Recognition," Collins, London, 1986.
2. J. T. Touand, R. C. Gonzalez. "Pattern Recognition Principles," Addison Wesley, 1974.
3. J. I. Fujimoto, T. Nakatani and M. Yoneyama. "Speaker-independent word recognition using fuzzy pattern mathing," Fuzzy Sets and Systems, 32. pp. 181-191, 1989.
4. W. K. Choi, H. H. Lee and K. Akizuki. "A vowel recognition by fuzzy inference and application to recognition of continuous korean speech," 電學論 C, Vol.III, No.5, 1990.
5. D. K. Burton, J. E. Shore and J. T. Buk. "Isolated word speech recognition using multisection vector quantization codebooks," IEEE Trans. Acoust., Speech, Signal, Vol.ASSP-33, No.4, Aug., 1985.
6. R. M. Gray and A. H. Gray, JR. "Rate distortion speech coding with a minnum discrimination information distortion measure," IEEE Trans. Acoust., Speech, Signal, Processing, Vol.ASSP 24, Oct., 1976.
7. A. Kandel. "Fuzzy Techniques in Pattern Recognition," John Wiley & Sons, 1982.
8. L. Lamel, I. R. Rabiner and A. E. Rosenberg, J. G. Wilpon. "An improved and point detector for isolated word recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol.ASSP-29, No.4. pp.777-785, Aug., 1981.

## ▲崔承禧



1955년 8월 24일생

1974년 3월 ~ 1981년 2월 : 전북  
대학교 물리학과(학  
사)1982년 9월 ~ 1984년 8월 : 명지  
대학교 대학원 전자  
공학과(석사)1988년 3월 ~ 1992년 2월 : 명지  
대학교 대학원 전자  
공학과(박사)

1981년 3월 ~ 1991년 4월 : 명지여자 고등학교

1992년 3월 ~ 현재 : 동신대학교 정보통신공학과

※주관심분야 : 디지털 신호처리, 음성인식 및 합성

## ▲崔甲石

현재 : 명지대학교 정보통신공학과 교수(제 11권 3호  
참조)