

# VQ와 Fuzzy 이론을 이용한 단어인식

## Word Recognition Using VQ and Fuzzy Theory

金子龍\* 崔甲石\*\*

(Ja Ryong KIM, Kap Seok CHOI)

### 요 약

음성인식에 있어서 문제점의 하나는 발성자에 따른 주파수 변동문제이다. 본 논문에서는 이러한 음성 신호의 주파수 특성의 변동에 따른 영향률 해결하기 위하여 fuzzy 이론을 도입하였다. 여기서 표준패턴은 음성신호의 대표적인 특징들을 포함하고 있어야 하므로, 먼저 여러 화자가 발성한 단어들을 벡터 양자화한 코드북을 생성하였으며, 이 코드북으로부터 추출한 피크 주파수와 피크 에너지를 fuzzy화 패턴으로 작성하였다.

입력 음성신호로부터 추출한 특징량인 스펙트럼의 피크 주파수와 피크에너지를 각각 멤버쉽 함수로 표현하여 fuzzy 추론에 의한 단어인식을 하였다. 실제 확신도 계산에 있어서는 계산량을 줄이기 위하여 fuzzy 값의 차만으로 확신도를 구하는 개선 확신도를 제안하여 사용하였다.

한국어 숫자음을 인식 실험한 결과 주파수 특성의 변동에 따른 영향률 해결할 수 있음을 확인하였으며, 제안된 개선 확신도 계산방법에 의해서 기억용량과 계산량을 감소시킬 수 있었다.

### ABSTRACT

The frequency variation among speakers is one of problems in the speech recognition.

This paper applies fuzzy theory to solve the variation problem of frequency features. Reference patterns are expressed by fuzzified patterns which are produced by the peak frequency and the peak energy extracted from codebooks which are generated from training words uttered by several speakers, as they should include common features of speech signals. Words are recognized by fuzzy inference which uses the certainty factor between the reference patterns and the test fuzzified patterns which are produced by the peak frequency and the peak energy extracted from the power spectrum of input speech signals. Practically, in computing the certainty factor, to reduce memory capacity and computation requirements we propose a new equation which calculates the improved certainty factor using only the difference between two fuzzy values.

As a result of experiments to test this word recognition method by fuzzy inference with Korean digits, it is shown that this word recognition method using the new equation presented in this paper, can solve the variation problem of frequency features and that the memory capacity and computation requirements are reduced.

### I. 서 론

\* 진주전문대학 전자계산과  
\*\* 명지대학교 전자공학과

으로 가장 자유로운 형태인 음성에 대해 많은 관심을 보여 왔으며, 최근 기계적인 도분없이 인간의 청도를 바로 수신으로 나타낼 수 있다는 인공성애 임적이라 사람의 기계와의 통선에 관한 여러가지 연구와 실험이 행해져 왔다. 그러나, 음성은 단순한 음향적 부호가 아니라 다양한 구성요소들의 언어로서의 규칙에 따라 구성되기 때문에 음향적 성질을 하나의 기호로 표현할 때 동일한 소리라도 지닌 의미와 발음 속도가 다를 수 있으며, 인간의 발성기관에 따라 복잡하게 변화하므로 음성인식의 연구에는 많은 어려움이 있다. 따라서 음성인식에 있어서 특징으로 추출되는 특징량들이 음성의 특징을 나타내는 절대적인 것으로 볼 수 없으므로 그 특징량들의 애매성을 표시해 줄 필요가 있다. 여기서 애매성을 표현하는 방법으로 1965년 Zadeh에 의해 제안된 fuzzy 개념은 확률적인 개념에서 완전히 탈피한 것으로 임의의 대상에 대한 불확실함을 fuzzy 집합으로 표현하여 대상의 불확실성을 판단하는 것이다.<sup>1)</sup>

음성인식에 대한 문제점의 하나로서 발성자에 따른 주파수 변동 문제가 있다. 본 논문에서는 이러한 음성신호의 주파수 특성의 변동에 따른 영향을 해결하기 위하여 fuzzy 이론을 도입하였다. 여기서 표준패턴은 음성신호의 대표적인 특징들을 포함하고 있어야 하므로, 먼저 여러 화자가 발성한 단어들을 벡터 양자화한 코드북을 생성하였으며, 이 코드북으로 부터 추출한 피크 주파수와 피크 에너지를 fuzzy화 패턴으로 작성하였다.

입력 음성신호로 부터 추출한 특징량인 스펙트럼의 피크 주파수와 피크에너지를 각각 멤버십 함수로 표현하여 fuzzy화 패턴을 작성하였고, 이 패턴들 사이의 확신도를 구하여 단어를 인식하는 fuzzy 추론에 의한 단어인식 방법을 제시하였다. 여기서 확신도는 fuzzy 연산 max-min에 의해 구해지며 확신도 계산을 위해서는 각 fuzzy 값에 대한 멤버십 값을 모두 기억해야 하기 때문에 많은 기억용량과 계산량이 필요하게 된다. 본 논문에서는 fuzzy 값의 차를 이용하여 확신도를 구하는 개선 확신도 계산식을 제안하여 기억용량과 계산량을 감소시켰다.

## II. Fuzzy 추론에 의한 단어인식

대부분의 패턴 인식 방법들은 직결을 기본으로 하는 통계적인 기법들을 사용하여 왔다. 그러나, 어떤 파라메타적인 확률 밀도 함수의 도입에 대한 임의의 선제는 정확하다고 볼 수는 없다. 즉 우리가 추정하는 특성들은 실제로 애매한 것들이기 때문에, 임의의 추정으로 실제의 분포를 알아내기 위하여 확률 밀도에 의한 통계적인 수법만을 사용한다는 것은 정확하다고 할 수 없다. 따라서, 임의의 독특한 함수를 선택보 하지 않는 방법의 도입<sup>2,3)</sup>이 요구된다.

### 2-1. 벡터 양자화

음성은 선형예측방법에 의하여 분석하면 피치, 에너지, 반사계수 그리고 예측계수 등을 구할 수 있다. 이 중에서 반사계수 또는 예측계수는 차지하는 데이터의 양이 많기 때문에, 최근에는 신호의 부호화로 많이 쓰이고 있는 데이터 압축기법인 벡터 양자화 방법을 이용하여 이들 정보를 압축하므로써 계산량을 감축시키고 있다. 따라서 본 논문에서는 음성신호가 가지고 있는 특징들을 모두 포함하거나 대표적인 특징으로 부호화해 주는 방법인 벡터 양자화로 작성된 코드북을 표준패턴으로 하였다.

### 2-2. Fuzzy 집합

전체 집합 U에 대한 fuzzy 집합(fuzzy set) A는 U의 각 요소  $\mu$ 가 A에 속한 정도를 나타내는 멤버십 함수  $\mu_A$ 에 의해 특성 지워진 멤버십들의 집합이다. 여기서  $\mu_A$ 는

$$\mu_A : U \rightarrow [0, 1] \quad (1)$$

로서, U의 각 요소에 0에서 1까지의 실수값  $\mu_A(u)$  ( $\in [0,1]$ )으로 주어지게 된다. 여기에서 멤버십 값이 1에 가까울 수록 요소  $\mu$ 는 A에 속한 정도가 크고, 0에 가까울 수록 A에 속한 정도가 작다.

### 2-3. Fuzzy 관계

Fuzzy 관계(relation) <sup>1)</sup>는 fuzzy 추론에서의 fuzzy 조건문을 표현하는 중요한 수법이다. 집합 A로부터 집합 B로의 fuzzy 관계 R이란 직積  $A \times B = \{(a,b) | a \in A, b \in B\}$ 에 있어서의 fuzzy 집합이며

$$\mu_R : A \times B \rightarrow [0, 1] \quad (2)$$

인 멤버쉽 함수  $\mu_R$ 로 특성지워 진다.

### 2-4. Fuzzy 추론

Fuzzy 추론은 몇개의 fuzzy 명제로 부터 어떤 하나의 명제를 도출하는 추론법이다. 가장 간단한 형식으로 식(3)과 같이 표현할 수 있다.

$$\begin{array}{l} \text{규칙 : } A \rightarrow B \\ \text{사실 : } A' \\ \hline \text{결론 : } B' \end{array} \quad (3)$$

이 fuzzy 추론의 결론 B'를 구하는 방법으로는 여러가지가 있으나, 여기서는 Zadeh의 fuzzy 추론법을 이용한다.

식(3)의 규칙  $A \rightarrow B$ 는 fuzzy 조건문이며, 그 멤버쉽 함수는

$$\mu_{A \rightarrow B}(u,v) = \mu_A(u) \rightarrow \mu_B(v) \quad (4)$$

로 정의된다. 결론 B'는 A'와  $A \rightarrow B$ 와의 max-min 합성 " "을 행함으로써 얻어지며, 이것을 추론의 합성규칙(composition rule of inference) <sup>2)</sup> 이라고 한다. 즉,

$$B' = A' \cdot (A \rightarrow B) \quad (5)$$

$$\mu_{B'}(v) = \bigvee_u \{ \mu_{A'}(u) \wedge [ \mu_A(u) \rightarrow \mu_B(v) ] \} \quad (6)$$

식(3)을 정의하는  $A \rightarrow B$ 는 많이 존재하나 여기서는 Mandani의 방법인

$$A \rightarrow B = A \wedge B \quad (7)$$

을 사용한다. 여기서 Mandani 방법인 식(7)을 식(6)에 적용하면 결론 B'는 다음 식과 같다.

$$\mu_{B'}(v) = \bigvee_u \{ \mu_{A'}(u) \wedge [ \mu_A(u) ] \wedge \mu_B(v) \} \quad (8)$$

여기서

$$h \equiv \bigvee_u \{ \mu_{A'}(u) \wedge [ \mu_A(u) ] \} \quad (9)$$

식(9)에서  $\bigvee$ 는 max를  $\wedge$ 는 min을 표시하며, Fuzzy 집합 A'와 A의 교집합의 최대값으로서 A'와 A의 '적합도'를 표시한다. 본 논문에서는 확신도로 사용하였으며, 확신도를 0과 1 사이의 수치로 표현하여, max-min 합성규칙으로 얻어지는 적합도를 구하는 연산에 의하여 확신도를 구한다.

#### 2-4-1. 생성규칙

음성을 Fuzzy 추론에 의해 인식을 행할 때에는 인식을 위한 생성규칙을 필요로 한다. 이 생성규칙은 전체에 따른 결론의 형태로서 다음과 같이 주어진다.<sup>4)</sup>

IF 전 제 THEN 결론

본 논문에서의 생성규칙은 음성신호에 대한 스펙트럼의 피크 주파수와 피크 에너지를 특징량으로 사용하므로 다음과 같이 생성규칙을 정하였다.

<p>IF</p> <p>음성신호의 피크 주파수 <math>F_i</math>가 fuzzy 값 <math>X^i</math>를 갖고,          음성신호의 피크 에너지 <math>E_i</math>가 fuzzy 값 <math>E^i</math>를 갖는다.</p> <p style="text-align: right;">(여기서, <math>i=1, 2, 3, 4</math>)</p> <p>THEN</p> <p>음성신호는 단어 "r"이다.</p>
--

#### 2-4-2. 입력패턴의 fuzzy화

음성인식을 위해 구한 특징들은 그 음성을 나타내는 절대적인 것은 아니며, 이 특징량을 동일인이 발성한다고 해도 그때 마다 다르게 된다. 이러한



이와 같이 표현되는 fuzzy화 패턴을 좀 더 명확하게 나타내기 위해, fuzzy화 패턴을 fuzzy 값과 멤버십 값과의 관계로 나타낸다. 여기서 멤버십 값은 fuzzy화 패턴의 의미를 어느 정도 포함하고 있는가를 나타내는 것으로 1.0에서 0.0 사이의 값을 갖고, 멤버십 값이 1.0인 경우는 fuzzy화 패턴의 의미를 완전하게 포함하는 것이며, 0.0일 때에는 완전하게 포함하지 않는다는 것을 의미한다. 본 예에서는 0.0과 1.0 사이의 멤버십 값을 중심 fuzzy 값에서 하나씩 떨어지는 정도에 따라서 시험패턴의 제 1 피크 주파수에 대한 fuzzy 값 61과 60의 멤버십 함수의 멤버십 값을 표 2-4와 같다. 따라서 두 패턴 사이의 확신도  $S(k)$ 은 식(9)를 본 연구의 확신도 계산에 적합하도록 식(10)과 같이 표현하여  $\vee - \wedge$  ( $\vee$ : max-min)에 의해 구한다.

$$S(k) = \vee(\mu_{ref}^{fk} \wedge \mu_{test}^{fk}) \quad (10)$$

단,  $k=1,2,3,4$ ( $k$ :  $k$ 번째 피크)

식(10)은 표 2-4의 예에서 먼저 표준패턴의 멤버십 함수의  $\mu_{ref}^{fk}$ 와 시험패턴의 멤버십 함수  $\mu_{test}^{fk}$ 의  $\wedge$ (min)을 취한 다음 그 중에서  $\vee$ (max)를 취하는 것이다. 즉, 식(10)의 우변 괄호안의 fuzzy 연산  $\wedge$ 의 결과는 두 fuzzy 값에 대한 결과로서 표 2-4의  $S^f(1)$ 과 같다.

또한, 식(10)의 나머지 fuzzy 연산  $\vee$ 의 결과는 괄호안의  $\wedge$  연산 결과 중에서 최대를 취하는 것이므로 확신도  $S^f(1)$ 은 0.99가 된다. 이 확신도  $S(1)$ 은 표준패턴에 대한 시험패턴의 첫번째 피크 주파수에 대한 확신도이다. 이러한 과정을 본 논문에서는 피크 주파수와 피크 에너지를 동시에 고려하였으며, 식(10)로 부터 피크 주파수와 피크 에너지에 대하여 각각의 확신도에 대한 평균을 취한 식 (11)을 정의한다.

$$S(k) = \frac{\vee(\mu_{ref}^{fk} \wedge \mu_{test}^{fk}) + \vee(\mu_{ref}^{ek} \wedge \mu_{test}^{ek})}{2} \quad k=1,2,3,4(k=k\text{번째 피크}) \quad (11)$$

여기서

표 2-4. 규칙 1의 제1 피크 주파수에 대한 멤버십 값  
Table 2-4. Membership values of the 1st peak frequency of the rule 1.

fuzzy 값	표준패턴의 $\mu_{ref}^{fk}$	시험패턴의 $\mu_{test}^{fk}$	$S^f(1)$
0	0.39	0.40	0.39
1	0.40	0.41	0.40
.	.	.	.
.	.	.	.
.	.	.	.
57	0.96	0.97	0.96
58	0.97	0.98	0.97
59	0.98	0.99	0.98
60	0.99	1.0	0.99
61	1.0	0.99	0.99
62	0.99	0.98	0.98
63	0.98	0.97	0.97
64	0.97	0.96	0.96
.	.	.	.
.	.	.	.
.	.	.	.
150	0.02	0.01	0.01
161	0.01	0.0	0.0
162	0.0	0.0	0.0
.	.	.	.
.	.	.	.
.	.	.	.
999	0.0	0.0	0.0

$\mu_{ref}^{fk}$ : 표준패턴인 피크 주파수에 대한  $k$ 번째 fuzzy 값의 멤버십 값들

$\mu_{test}^{fk}$ : 시험패턴인 피크 주파수에 대한  $k$ 번째 fuzzy 값의 멤버십 값들

$\mu_{ref}^{ek}$ : 표준패턴인 피크 주파수에 대한  $k$ 번째 fuzzy 값의 멤버십 값들

$\mu_{test}^{ek}$ :  $k$ 번째 피크 주파수와 피크 에너지에 대한 확신도 값

### III. 인식 알고리즘

#### 3-1 확신도의 측정

단어 인식을 위한 전처리 과정에서는 그 단어에 대한 특징을 추출한다. 본 연구에서는 단어 음성의 프레임별 LPC 계수를 구하고, 그것으로부터 각 프레임별 피크 주파수와 그의 에너지를 특징량으로 추출하였다. 이때 같은 단어를 발성하더라도 발음속도와 발성 기관에 따라 특징량이 다르게 나타날

수 있다. 이러한 음성의 변동분을 인식과정에서 fuzzy set 개념을 도입해 줌으로써 음성 신호의 불확실한 성분도 모호한 불확실성을 나타내는 fuzzy set의 대응시켜 fuzzy화 패턴으로 변환된 fuzzy 값을 이용하여 fuzzy 추론에 의한 인식을 행하였다.

시험패턴이 표준패턴에 어느 정도 유사하게 되어 있는가를 관찰하기 위해서는, 시험패턴에 대한 평가 규칙의 전체가 가동되고 시험패턴의 fuzzy 화 패턴이 그 전체를 어느 정도 만족하고 있는가를 측정하기 위하여 불확실한 데이터에 대한 확신도를 구해야 한다. 본 논문에서는 피크 주파수와 피크 에너지에 대한 확신도, 프레임별 확신도, 코드워드 선택 확신도 및 최종 확신도를 각각 정의하였으며, 그것을 차례로 계산하여 단어 인식에 사용하였다.

(1) 피크 주파수와 피크 에너지에 대한 확신도 : 식(10)은 피크 주파수의 표준패턴과 시험패턴 사이의 확신도를 나타낸다. 본 논문에서는 피크 주파수의 피크 에너지를 동시에 고려한 확신도를 각각의 확신도에 대한 평균을 취한 식(11)로 정의하였으며, 실제의 계산에 있어서는 기억용량을 줄이기 위해서 개선 확신도 계산식(15)을 사용하였다.

(2) 프레임별 확신도 : 각 프레임의 특징량인 피크 주파수와 피크 에너지는 4개로 하였다. 그러나, 두 프레임 사이의 확신도를 구하기 위해서는 프레임별 특징량인 4개의 피크에 대해 식(15)와 같이 확신도를 구해야 하며, 이 4개의 확신도로 부터 프레임과 프레임 사이의 유사성을 나타내는 프레임의 확신도를 정해야 한다. 본 논문에서는 프레임의 확신도를 4개의 확신도 중에서 최소인 것으로 하였으며 다음식과 같이 표시하였다.

$$SIM(m) = \min_K \{S(k)\} \quad (12)$$

(3) 코드워드 선택 확신도 : 프레임 확신도는 표준패턴의 각 코드워드와 시험패턴의 프레임 사이에서 측정하게 되므로 어떠한 코드워드가 시험패턴과 가장 유사한지 선택해야 한다. 이를 위해 프레임 확신도중에서 최대인 것을 선택 하였으며, 다음과 같이 표시하였다.

$$C(m) = C(m) + \max_n \{SIM(m)\} \quad (13)$$

이와 같은 식(12)과 식(13)의 결과패턴들의 최종 확신도, 최종 코드워드, 최종 프레임, 최종 패턴, 시험패턴이나 유사한가는 코드워드 선택 확신도들이 얼마나 큰지에 따라 결정된다. 그런데, 시험패턴은 단어이기 때문에 각 프레임 수가 불규칙하므로, 각 프레임별로 구한 코드워드 선택 확신도를 누적하고 프레임 수로 나누어 줌으로써 표준패턴에 대한 시험패턴의 최종 확신도를 구하였다. 최종 확신도를 다음식과 같이 표시하였다.

$$CER(i) = C(i) \quad (14)$$

### 3-2. 개선 확신도

식(11)에 의한 확신도를 계산할 경우 fuzzy 값에 대한 멤버십 값을 모두 기억시켜야 하므로 많은 기억 용량이 소요된다. 따라서 본 논문에서는 멤버십 값을 모두 기억시키지 않고 각 특징량에 대한 fuzzy 값의 차를 이용하여 확신도를 구할 수 있는 새로운 개선 확신도를 식(15)으로 정의한다.

$$S(k) = \frac{g \left\{ \frac{M^f x - (X^f \text{ref} - X^f \text{test}) / 2}{M^f x} \right\} \cdot g \left\{ \frac{M^e x - (X^e \text{ref} - X^e \text{test}) / 2}{M^e x} \right\}}{2} \quad (15)$$

여기서,

$$g(x) = \begin{cases} x & : 0 \leq x \leq 1 \\ 0 & : x < 0 \end{cases}$$

$X^{fk}_{ref}$  : 표준패턴인 피크 주파수에 대한 k번째 fuzzy 값

$X^{fk}_{test}$  : 시험패턴인 피크 주파수에 대한 k번째 fuzzy 값

$X^{ek}_{ref}$  : 표준패턴인 피크 에너지에 대한 k번째 fuzzy 값

$X^{ek}_{test}$  : 시험패턴인 피크 에너지에 대한 k번째 fuzzy 값

$S(k)$  : k번째 피크 주파수와 피크 에너지에 대한

확신도 값

$M^f_x$  : 피크 주파수에 대한 삼각 멤버십함수의 fuzzy 값 범위의 1/2

$M^e_x$  : 피크 에너지에 대한 삼각 멤버십함수의 fuzzy 값 범위의 1/2

본 논문에서는 피크 주파수와 피크 에너지에 대한 삼각 멤버십 함수의 fuzzy 값의 범위  $M^f_x$ ,  $M^e_x$ 는 주정량의 변동을 유사해 주는 정도로서 그 범위에 따라 흡수되는 정도의 차가 생기므로 적당한 값을 선정해 주어야 한다.

### 3-3. 인식 알고리즘

위의 과정에 대한 인식 알고리즘은 그림 3-1과 같다.

단어를 인식하는 과정은 다음과 같다. 시험패턴의 첫번째 프레임 스펙트럼의 제 1 피크 주파수와 피크 에너지에 대하여 표준 패턴의 첫번째 코드워드의

스펙트럼에 대한 세 1 피크 주파수와 피크 에너지의 확신도를 식(15)와 같이 구한다. 이 확신도를 제 2, 제 3, 제 4 피크 주파수와 피크 에너지에 대하여 반복하여 구한 후, 그 중에서 식(12)와 같이 확신도가 최소인 것을 선택하므로써 프레임 확신도를 구하게 된다. 이러한 과정을 모든 코드워드에 대해서 구한 후 어떠한 코드워드가 시험패턴의 첫번째 프레임과 가장 유사한지 선택하기 위하여 식(13)과 같이 코드워드의 선택 확신도를 구하게 된다.

이상의 과정으로 전 프레임에 대해서 각각의 코드워드의 선택 확신도를 구하여 누적한 후, 프레임 수로 나누어 줌으로써, 시험패턴에 대한 최종 확신도를 식(14)와 같이 구하게 된다. 표준패턴인 모든 코드북에 대해서 같은 과정을 통해 얻어진 각각의 최종 확신도 중에서 확신도가 최대로 되는 단어를 인식단어로 한다.

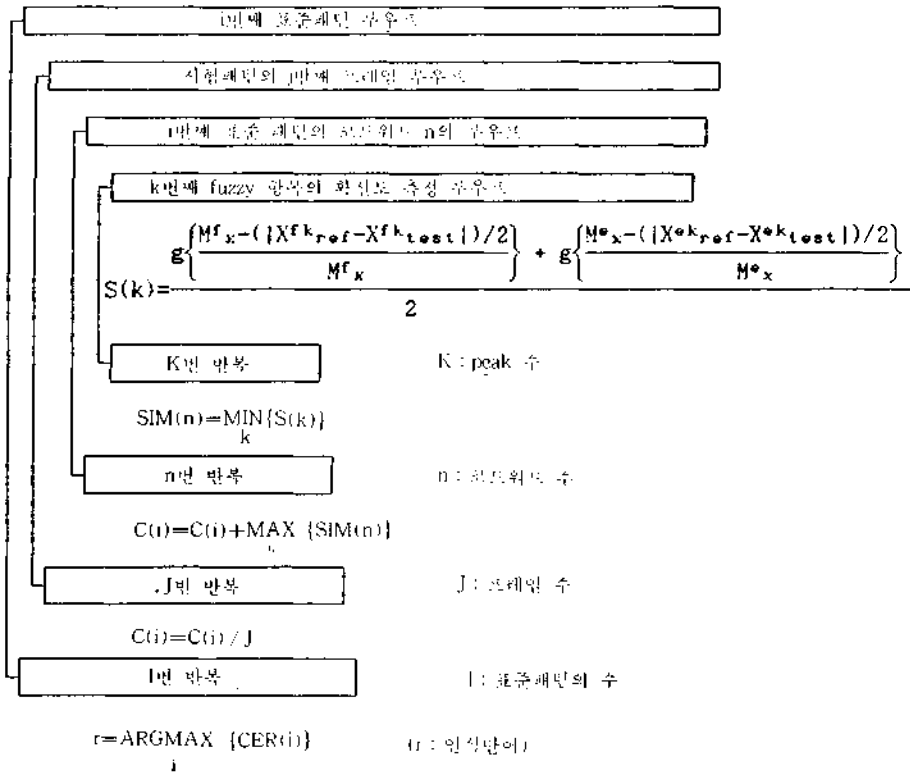


그림 3-1. Fuzzy 추론에 의한 단어인식 알고리즘  
Fig. 3-1. Word recognition algorithm using fuzzy inference.

### IV. 실험 및 고찰

#### 4-1. Fuzzy화 패턴작성

먼저 fuzzy화 패턴 작성을 하기 전에 단어인식 시스템에 음성이 입력되면 음성신호만을 효과적으로 처리해 주기 위해 신호의 크기와 잉코차율을 이용하여 음성신호의 시작점과 끝점을 검출하였다. 이상과 같이 검출된 음성데이터를 분석하기 위한 선형예측 분석 알고리즘으로는 최소자승법을 사용하였으며, 각 프레임의 길이는 256(25.6[ms])개로 하였고, 각 프레임의 이동구간은 100(10[ms])개로 하였다.

본 논문에서 제시한 벡터 양자화와 fuzzy 추론을 이용한 단어인식 방법의 성능을 평가하기 위하여 한국어 숫자음(0,1,2,...,9)을 음성 데이터로 하여 인식실험을 수행하였으며, 이 음성 데이터는 7인의 화자가 각각 5회씩 발성하였으므로, 실험에서 사용한 총 음성 데이터 수는 7(인)×5(회)×10(개)=350(개)이다. 여기서 7인이 2회씩 발성한 음성 데이터 중 7(인)×2(회)×10(개)=140(개)는 표준패턴으로 나머지 7(인)×3(회)×10(개)=210(개)는 시험패턴으로 사용하였다.

이상의 음성 데이터로 패턴을 작성하기 위한 선처리 작업으로, 먼저 음성데이터의 끝점을 검출하고 선형예측 분석을 하였다. 그 분석 조건은 표 4-1과 같다.

표준 패턴은 단어별로 벡터 양자화에 의해 생성한 코드북의 코드워드에서 추출한 피크 주파수와 피크

표 4-1. 분석 조건

Table 4-1. Analysis Conditions

LPC Sampling Freq.	4KHz
A/D Data	10kHz Sampling, 12 bit
Window Function	Hamming Window and Pre-emphasis(A=0.95)
Frame Length	20ms
Frame Period	10ms
Analysis	14 order LPC
Distortion Measure	Likelihood Ratio

에너지를 fuzzy 값으로 표현하여 작성하였으며, 시험 패턴은 음성 데이터의 선형예측 분석에 의한 파워 스펙트럼에서 피크 주파수와 피크 에너지를 추출하여 fuzzy 값으로 표현하여 작성하였다. 이와 같은 패턴을 작성하기 위한 구성도는 그림 4-1과 같다.

#### 4-2. 멤버쉽함수에 대한 fuzzy 값의 범위

Fuzzy 이론에서 확신도는 fuzzy 연산 max-min에 의해 구해지며, 확신도 계산을 위해서는 fuzzy 값에 대한 멤버쉽 값을 모두 기억해야 하기 때문에 기억용량에 대해서 고려할 필요가 있다. 그러나 본 논문에서 제시한 식(15)를 이용하면 각 fuzzy 값에 대한 멤버쉽 값을 모두 기억하지 않고도 각 특징량에 대한 fuzzy 값의 차만을 이용하여 확신도 값을 구할 수 있다. 식(15)에 의해서 확신도를 구할 경우 삼각 멤버쉽 함수의 fuzzy 값 범위를 어떻게 설정하느냐에 따라 동일한 패턴이라도 그림 4-2와 같이 확신도 값이 달라진다. 따라서 가장 좋은 확신도를 얻기 위하여 fuzzy 값 범위들 가장 적절한 범위로

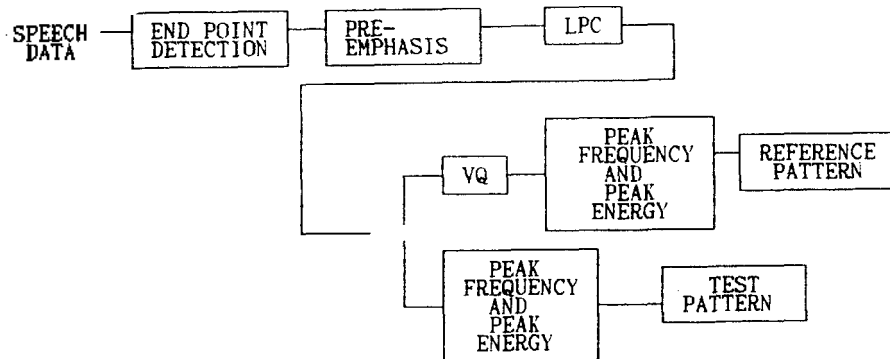


그림 4-1. Fuzzy화 패턴 작성 시스템 구성도  
Fig. 4-1. Block diagram for system of fuzzified pattern.



선정된 필요가 있다. 본 논문에서는 실험을 통하여 피크 주파수와 피크 에너지에 대한 fuzzy 값의 범위를 설정하였다.

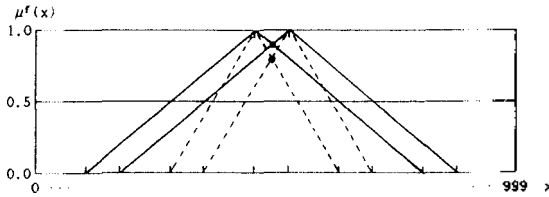


그림 4-2. Fuzzy 값의 범위에 따른 확신도의 변화  
Fig. 4-2. Variation of certainty factor according to range of fuzzified values.

4-3. 결과 및 고찰

본 논문에서는 선형예측 분석과 벡터 양자화를 이용하여 전처리 한 다음 fuzzy 추론을 적용하여 단어 인식을 하였다. 여기에서는 같은 전처리 과정에서 얻어진 데이터를 인식 방법에 따라 그 결과를 비교한다.

그림 4-3은 피크 주파수에 대한 인식 방법으로 Euclidean 거리를 이용하여 인식한 결과, 피크 주파수에 대한 인식방법으로 fuzzy 추론을 적용하였을 경우의 인식 결과 및 피크 주파수와 피크 에너지에 대해 fuzzy 추론을 적용하였을 경우의 인식율이다.

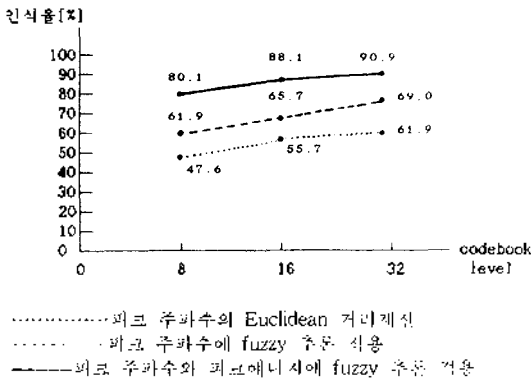


그림 4-3. 인식방법과 코드워드 수에 따른 인식율 비교  
Fig. 4-3. Comparison of recognition rate according to recognition methods and codewords.

그림 4-3을 보면, 피크 주파수의 Euclidean 거리와 이용한 방법에서는 코드워드의 수를 8, 16, 32로

높아짐에 따라 47.6[%], 55.7[%], 61.9[%]로 평균 55.0[%]의 인식율을 얻었고, 피크 에너지에 fuzzy 추론을 적용한 경우에는 각각 61.9[%], 65.7[%], 69.0[%]로 평균 65.5[%]의 인식율을 얻었다. 본 논문에서 제시한 벡터 양자화와 fuzzy 추론을 이용한 방법에서의 인식율은 각각 80.0[%], 88.1[%], 90.9[%]로써 평균 86.0[%]의 인식율을 얻었다. 따라서 피크 주파수만의 Euclidean 거리를 이용한 인식율에 비해 fuzzy 추론을 적용하였을 경우 약 10[%] 정도의 인식율 향상이 있었으며, 피크 주파수와 피크 에너지에 fuzzy 추론을 적용하였을 경우에는 약 30[%]의 인식율 향상이 있었다. 또한 피크 주파수의 Euclidean 거리를 이용한 방법에서 인식율이 낮은 것은 단순히 피크 주파수끼리의 거리 계산만으로 음성신호의 특성을 파악해 내기는 어려운 것이라 생각된다. 이에 대하여 본 논문에서 제시한 벡터 양자화와 fuzzy 이론을 이용한 단어인식 방법에서는 피크 주파수와 피크 에너지로 먼저 fuzzy화 패턴을 작성하고 식(18)을 이용하여 확신도를 구하기 때문에 주파수 변동에 대한 영향을 흡수할 수 있었으며 인식율 향상이 있었다. 즉, 주파수 변동에 대한 영향의 흡수분제는 각 피크 주파수가 다소 변하더라도 멤버십 함수에 적용하면 유사성이 높을수록 높은 확신도를 갖게되므로 주파수 변동에 따른 영향을 해결할 수 있다.

V. 결 론

본 논문에서는 음성신호의 주파수특성의 변동에 따른 영향을 해결하기 위해 음성신호 부터 추출한 피크 주파수와 피크 에너지를 fuzzy화 패턴으로 작성하였다. 이러한 fuzzy화 패턴을 fuzzy 추론에 적용하여 단어를 인식한 결과 다음과 같은 결론을 얻을 수 있었다.

첫째, fuzzy 추론에 의한 단어인식 방법에서 주파수 특성을 멤버십 함수로 표현해 줌으로써 주파수 특성의 변동에 따른 영향을 해결할 수 있었다.

이를 평가하기 위하여 피크 주파수에 Euclidean 거리를 적용하여 단어를 인식한 경우와 fuzzy 추론을 적용한 경우를 비교한 결과 10[%] 이상의 인식

을 향상이 있었으며, 피크 에너지를 추가하여 fuzzy 추론을 적용한 경우에는 30[%] 이상의 인식률 향상이 있음으로 주파수 변동에 따른 영향이 유수됨을 알 수 있었다.

둘째, 두 패턴의 fuzzy 값에 대한 차를 이용한 개선 확신도 계산식을 제안하여 사용하므로써, max-min 연산에 의한 확신도의 경우보다 기억용량을 약 1/500로 감소시킬 수 있었으며, 단어인식에 있어서 fuzzy 추론은 주파수 특성의 변동이 많은 음성신호에 대해 효과적인 방법임을 알 수 있었다.

더불어 본 연구의 방법이 불특정 화자가 발성한 단어인식 실험에도 성과가 있을 것으로 생각된다.

#### ▲김 자 룡 10권 2호 참조

#### 參 考 文 獻

1. 廣田薫, "フジィ 情報處理應用の現狀と展望", 情報處理 Vol.30, No.8, pp. 913-921, Aug. 1989.
2. 藤本潤一郎, "フアジィ 理論を用いた音聲認識", 情報處理, Vol.30, No.8, pp. 957-962, Aug. 1989.
3. J.I. Fujimoto and T. Nakatani and M. Yoneyama, "Speaker-independent word recognition using fuzzy pattern matching", Fuzzy Sets and System 32, pp. 181-191, 1989.
4. 藤本潤一郎 "フアジィ 音聲認識", 數理科學, No.284, pp.63-69, Feb. 1987.
5. 寺野 壽郎, 淺居 喜代治, 菅野 道夫, "Fuzzy system theory and its application", オ-ヌム 社, pp. 45-66, 1989.
6. 林陽一, 中井正人, "Automated extraction of fuzzy IF THEN rules using neural networks", 電學論C, 1 10卷 3號, 平成2年, pp. 198-206.

#### ▲최 갑 석(Kap Seok Choi) 1930년 9 월30일생



1951년 9 월~1955년 3 월 :

서울대학교공과대  
학 통신공학과 졸업

1975년 2 월25일 : 명지대학교  
공학박사

1980년 8 월~1981년 8 월 :

일본 동경대학 공  
학부 전자공학과  
연수

1972년 3 월~현재 : 명지대학교 전자공학과 교수