

# Fuzzy를 이용한 VQ/NN에 기초를 둔 음성 인식

## Speech Recognition Based on VQ/NN using Fuzzy

안 태 옥\*  
(Tae Ock Ann\*)

### 요 약

본 논문은 불특정 화자의 단모음 인식에 관한 연구로써, fuzzy개념을 이용한 VQ(Vector Quantization)/NN(Neural Network)에 의한 음성 인식 방법을 제안한다.

이 방법은 fuzzy를 이용하여 VQ codebook에 의해 다중 관측열(multi-observation sequence)을 구해 각 symbol이 데이터로부터 가질 수 있는 확률값을 계산하여 이 값을 신경 회로망의 입력으로 사용하는 방법이다.

인식 대상어로는 한국어 단모음을 선정하였으며 10명의 남성 화자가 8개의 단모음을 10번씩 발음한 음성 데이터베이스를 이용하여 fuzzy를 이용하지 않은 VQ/NN과 fuzzy를 이용한 VQ/HMM(hidden Markov model)에 의한 인식률과 비교 실험한다.

실험 결과에 의하면, VQ/NN에 의한 인식률은 92.3%이며, fuzzy를 이용한 VQ/HMM에 의한 인식률은 93.8%이고, fuzzy를 이용한 VQ/NN에 의한 인식률은 95.7%이다.

그러므로, 본 연구의 fuzzy를 이용한 VQ/NN이 학습 능력이 뛰어난 관계로 fuzzy를 이용한 VQ/HMM과 일반적인 VQ/NN 보다 인식률이 향상됨을 보여준다.

### ABSTRACT

This paper is the study for recognizing single vowels of speaker-independent, and we suppose a method of speech recognition using VQ(Vector Quantization)/NN(Neural Network).

This method makes a VQ codebook, which is used for obtaining the observation sequence, and then calculates the probability value by comparing each codeword with the data, finally uses these probability values for the input value of the neural network.

Korean single vowels are selected for our recognition experiment, and ten male speakers pronounced eight single vowels ten times. We compare the performance of our method with those of fuzzy VQ/HMM and conventional VQ/NN

According to the experiment result, the recognition rate by VQ/NN is 92.3%, by VQ/HMM using fuzzy is 93.8% and by VQ/NN using fuzzy is 95.7%.

Therefore, it is shown that recognition rate of speech recognition by fuzzy VQ/NN is better than those of fuzzy VQ/HMM and conventional VQ/HMM because of its excellent learning ability.

### I. 소 개

음성은 인간의 가장 자연스러운 통신 방법으로, 인간과 기계 사이의 자연스러운 통신을 위해 음성 인식에 대한 연구가 꾸준히 진행되어 왔다. 사람의 음성을 인식하기 위한 알고리즘으로는 그 동안 DTW(dynamic time warping)<sup>1)</sup>, VQ(vector quantization)<sup>2)</sup>, HMM(hidden Markov model)<sup>3, 4)</sup> 등이 이용되었는데 그 중에서도 HMM은 확률

론을 도입한 알고리즘으로 1980년대 이후에 다른 방법보다 우수한 장점이 있는 관계로 많은 연구가 이루어지고 있다. 그러나, 이 HMM도 연속 음성, 무한대어휘, 화자 독립이라는 음성 인식의 세가지 최종 목표를 해결하는데 여러 문제점이 지적됨에 따라 새로운 방법에 의한 음성 인식 방법이 대두되었는데 그것이 바로 fuzzy와 인공 신경 회로망을 이용한 음성 인식에 관한 연구이다.

이 fuzzy와 신경 회로망이란 둘 다 모두 인간이 생각하는 방법과 비슷하게 처리할 수 있다는 사실에 근거하여 새로운 계산 방식 및 패턴 인식의 방법으로 음성 인식에 적용되어 연구되고 있다.

\*전북산업대학교 컴퓨터공학과  
접수일자: 1995년 9월 18일

따라서, 본 연구에서는 새로운 방법의 fuzzy를 이용한 VQ/NN에 의한 음성 인식 방법을 제안한다. 이 fuzzy를 이용한 VQ/NN에 의한 음성 인식 방법은 이산 HMM에서 VQ codebook으로부터 fuzzy rule를 이용하여 다중관측열을 구하고 이를 이용하여 VQ/HMM 파라메타(parameter)들을 최적화 시키는 것과는 달리, fuzzy를 이용한 HMM에서처럼 VQ codebook의 codeword와 각 토큰(token)의 프레임과의 거리값을 계산하여 fuzzy를 이용해서 다중관측열을 구한 후 각각의 codeword가 가질 확률값을 계산하고 이를 입력 특징으로 하였으며 학습 알고리즘으로는 Rumelhart등이 제안한 back propagation 알고리즘을 이용하였다.

본 연구는 제안된 fuzzy를 이용한 VQ/Neural net에 의한 음성 인식 방법 이외에도 비교를 위해 최근까지도 많이 이용되고 있는 인식 방법인 VQ/HMM에 fuzzy rule을 적용시킨 fuzzy를 이용한 VQ/HMM에 의한 음성 인식과 일반적인 VQ/NN에 의한 음성 인식과 비교한다. 여기에서 VQ codebook을 작성하는데 사용된 vector는 10차 LPC cepstrum 계수이고, 대상 어휘는 8개의 단모음이며, 화자는 10명의 남성으로써 3명이 5번 발음한 음성으로 학습시켰으며, 7명이 발음한 560개의 단모음을 인식 실험하였다.

## II. 이 론

### 1. VQ(vector quantization) codebook 작성

VQ란 벡터의 sequence를 통신이나 디지털 채널에 저장 하기에 적당한 디지털 sequence와 mapping 하기 위한 시스템이다. VQ의 가장 큰 목적은 데이터 압축으로 데이터의 신뢰성을 잃지 않으며, 최대 한도로 bit rate를 줄이는데 있다. 데이터 압축에 기여한 Shannon의 rate distortion 이론<sup>7)</sup>에 의하면 스칼라 대신에 벡터를 코딩함으로써 더 좋은 성능을 얻을 수 있다는 것이다. 그러므로, 음성인식에 있어서 데이터 압축이라는 측면에서 VQ를 이용하였다. 즉, VQ는 입력 음성의 특징 벡터를 이미 저장되어 있는 특징 벡터들 중의 하나로 mapping시켜 주는 것을 의미한다.

그러므로, VQ를 음성인식에 이용하면 codebook이 reference template가 되므로 기억 용량을 작게 할 수 있고, 또한 codebook의 크기가 작음에 따라 인식하는데 걸리는 시간도 비교적 적게 걸린다. 그러나, VQ에 의한 음성 인식은 인식률이 떨어진다. 따라서, 최근에 VQ에 의한 음성 인식보다는 이 VQ를 이용한 음성 인식을 많이 하는데 그 방법 중의 하나가 VQ/HMM에 의한 음성 인식 방법이다.

이에 따라, 본 연구에서도 이 VQ codebook을 신경 회로망에 이용하는데 이 때 관측열을 구하는데 한 벡터에 하나의 심볼만 선택하던 종전의 방법과는 달리 한 벡터에 다중의 심볼을 선택하는 fuzzy 개념에 의해 VQ/NN하

여 일반적인 VQ/NN<sup>1)</sup>과 비교하고 VQ/HMM에 의한 음성 인식 방법과 비교하겠다. 학습 벡터들에 의해 codebook이 만들어지며, 벡터들은 codebook의 벡터들 중에서 거리값이 작은 순으로 fuzzy rule에 의해 다중의 심볼을 선택하게 된다.

본 연구에서는 전 단어를 training data로 삼아 하나의 codebook을 취하는 방법을 사용하였다. 이 때 codeword를 구하는 방법은 clustering 기법 중 K-means 알고리즘<sup>8)</sup>을 사용하였으며, 중심점 잡은 방법으로는 averaging 기법을 이용하였고 단모음 8개를 대상으로 인식 실험을 행한 관계로 codebook의 크기는 32로 하였다.

여기서, Test set vector를  $C_i$ ,  $\{i=1, \dots, 1\}$ 로 하면, 이 벡터들은 어휘에서의 단어들이 다양한 화자에 의해서 발음될 때 일어나는 vector이다. VQ에 숨은 주요 개념은 주어진 codeword의 수 M에 대하여 가장 가까이에 있는 codebook entry  $C_m$ 에 의해 test set vector  $C_i$ 의 각각에 대해 평균 거리가 최소가 되도록 최적의 codebook의 집합을 결정하는 것이다. Vector  $C_i$ 와 codeword  $C_m$ 간의 거리를  $d(C_m, C_i)$ 라고 정의한다면, 그 때 VQ의 목표는 관측열을 찾는 것이다.

VQ에서 사용하는 국부적인 거리는 본 연구에서 사용하는 LPC cepstrum의 계수를 P차로 하였을 경우에 다음과 같다.

$$d(C_m, C_i) = 0.04 * (C_{m0} - C_{i0})^2 + \sum_{p=1}^P (C_{mp} - C_{ip})^2 \quad (1)$$

### 2. fuzzy 개념을 이용한 다중 관측열 나열

인간과 기계와의 대화로 맨·머신 인터페이스라는 말을 자주 들을 수 있다. 로봇과 컴퓨터와의 대화를 생각할 필요가 생긴 지금 말을 사용하여 친밀하게 대화를 하려고 생각하는 것은 매우 자연스런 것이다. 앞에서 서술한 fuzzy 이론의 개념을 음성 인식에 도입함으로써 화자의 차이와 시간적인 변동을 흡수 할 수 있다.

같은 음성에서도 발성 속도의 차이가 원인으로 음성의 길이가 일정하지 않다. 같은 사람이 같은 말을 하여도 그때마다 그 길이가 바뀌어진다. 물론 똑같은 말이라도 발성하는 사람이 다르며 길이의 변동 역시 크다. 또한 발성 기관의 크기는 인간에 따라 달라 같은 형태로 하여 발성 하여도 공진 주파수에 차이가 생긴다. 이것이 패턴상의 개인성이 되어 나타난다. 화자에 의한 변동은 우리들 인간이 누구의 목소리라도 인식 할 수 있기 때문에 용이하다고 생각되지만 실제로는 매우 어려운 문제이다.

그래서, 이를 극복하려는 여러 가지 기법들이 나왔다. 예를 들면, DTW의 경우는 멀티 템플레이트법으로 해결해 보려했고, 일반적인 HMM의 경우는 많은 화자의 많은 학습 데이터를 통계적인 처리를 행함으로써 이를 극복하려 했다. 그런데, 이런 노력에도 불구하고 이것을 완전히 해결하지 못함으로 본 연구에서는 fuzzy 이론을 이용하여 극복하려 한다. 어떤 발음을 했을 때 발음할 때

다 또는 화자에 따라 특성이 달라지지만 그 발음이 가지는 특성이 많이 벗어나지는 않을 거라는데 유의하여 본 연구에서는 fuzzy 이론을 이용한다.

즉, VQ 코드북의 각 코드워드와 각 프레임의 LPC 캡스트럼 벡터간의 거리를 구하는 식은 위의 식 (1)에서 볼 수 있는 바와 같으며 각 프레임의 벡터가 코드워드 중 거리값이 가장 짧은 것을 심볼로 선택하는 것이 일반적인 HMM에서 관측열을 구하는 방법이었다.

그런데, 본 연구에서는 일반적인 HMM에서 관측열을 구하는 것과는 달리 각 열마다 다중의 심볼을 선택하는 방법을 제안하였는데, 이 방법은 일반적인 HMM이 학습할 때 관측열로 선택되지 않은 벡터는 확률값이 0으로 떨어지는데 반해 그런 벡터의 경우도 몇개의 다중 심볼 안에 들면 확률값을 가질 수 있으므로 인식시에 심볼로 선택되며 확률을 가지므로 인식될 수 있다는 개념에서 비롯된다.

실험에서 사용된 다중관측열을 구하는데 사용한 fuzzy 법칙<sup>11)</sup>은 다음과 같다.

$$w_r = \frac{2^{R-r}}{\sum_{i=1}^R 2^{i-1}} \quad (2)$$

여기서, r는 어떤 프레임의 벡터와 코드북의 각 코드워드 중 거리값이 작은 순으로 표시했을 때 몇 번째인지를 나타내고, R는 한 열에서 선택될 다중 심볼의 수를 의미하며,  $w_r$ 는 전체 확률을 1로 했을 때 거리값이 r 번째인 심볼이 가질 확률값이다.

### 3. 제안된 fuzzy를 이용한 VQ/NN에 의한 음성 인식

신경 회로망은 인간의 두뇌의 생물학적 신경 계통에 근거한 간단하고 많은 처리 요소들을 병렬로 상호 연결하여 학습을 통해 입력 패턴에 내재해 있는 정보를 분산 및 병렬 처리하는 정보 처리 시스템으로, 기존의 음성 인식 알고리즘에서 해결할 수 없었던 패턴 인식 문제를 인공 신경 회로망을 이용하여 해결하려 한다.

본 연구에서 사용된 신경 회로망은 Back Propagation 학습 알고리즘을 사용하는 MLP 구조로 이 구조의 입력 갯수는 VQ 코드북의 크기와 같으며, 입력값은 VQ/HMM에서와는 달리 코드북의 각 Codeword가 어떤 데이터에 대해서 VQ의 열로서 선택될 확률값이다.

#### 3.1 VQ를 이용한 입력값 결정

본 연구의 Neural Net의 입력 특징 파라메타는 VQ 코드북을 이용하여 구하게 되는데, 앞서서도 언급한 바와 같이 신경 회로망으로는 MLP를 이용하는데 이때 입력 특징 파라메타의 수는 Codeword의 수와 같으며 따라서 Codebook의 크기를 32로 하여 실험한 관계로, 본 연구에서는 32개의 입력값을 가지게 된다. 여기서 입력값을 구

하는 방법은 식 (1)에 의해서 구해진  $C_m$ 의 집합을 이용한다. 즉, m번째 입력값을  $x_m$ 이라 했을 때 다음 식에 의해 각 Codeword에 의한 입력값이 결정된다.

$$x_m = \frac{1}{N} \sum_{n=1}^N \delta(C_n) \quad (3)$$

여기서,

$$\delta(C_n) = \begin{cases} 1 & \text{if } C_m = C_n \\ 0 & \text{if } C_m \neq C_n \end{cases}$$

이고, N은 VQ를 구성하는 Test Data의 Frame 수이다.

#### 3.2 학습 및 인식 방법

패턴 인식에 널리 쓰이는 multi-layered perceptron의 구조는 hidden layer를 포함하는 feedforward multi-layered network 구조를 갖는다. 본 연구에서 사용한 학습 알고리즘은 generalized Delta rule에 근거한 Rumelhart의 back propagation 알고리즘이다. 이 알고리즘에 의한 학습 절차는 크게 두 단계로 나눌 수 있는데, 그 첫단계는 인공 신경 회로망에 입력값을 제시하고 각 노드에 대해서 network의 입력함수와 활성화 함수를 이용해서 출력을 산출하는 forward pass이고, 두번째 단계는 desired output과 actual output과의 차이를 계산하여 이 차이를 back propagation 시키면서 layer와 layer 사이의 가중치를 조절하는 backward pass이다. 이 두 단계는 시스템이 안정 될 때까지 즉, total error sum이 error criterion을 넘지 않을 때까지 계속 반복 실행한다. 이런 개념에서 본 연구에서 이용한 back propagation 학습 알고리즘은 참고 논문 [9]에 따른다.

#### 3.3. 실험 조건<sup>12)</sup>

본 연구에서는 단모음 중 이중 모음화한 “기”와 “니”를 제외한 8개의 단모음을 대상으로 단모음 인식을 행하였는데 이중 “기”와 “니”를 구별해서 발음하는 화자들이 없는 관계로 “기”와 “니”는 같은 발음으로 간주해서 학습시에도 같은 class로 생각하고 학습시켰으며, 인식시에도 구별없이 같은 발음으로 간주하여 인식시켰다.

따라서, 8개의 단모음을 7 class로 간주하여 7개의 출력을 주어 학습 시켰는데 1000번 반복하여 학습된 후의 학습시의 예러는 20000번까지 반복해서 학습시켜도 개선되지 않았으며 또한 그 값이 0.05에서 0.09까지 진동하였으며 그 이상의 개선이 되지 않았다. 또한 인식시켰을 경우에도 대부분이 인식되지 않았다.

그 이유는 back propagation 알고리즘은 정적 특징을 학습시키는데는 적합하나 음성은 정적 특징과 함께 동적 특징도 있어 그 출력 node 수를 증가시킬 경우에도 특징간의 classification을 잡아내지 못하기 때문이다.

따라서, 본 연구에서는 NN의 병렬성은 고려하지 않고 HMM처럼 학습 능력만을 고려하여 실험하였다. 즉 8개의 단모음을 7개의 class로 간주하여 7개의 출력값을 고려한 것이 아니고, 학습 데이터를 가지고 각 단모음별로 그 단모음에 속하는지 않는지를 고려하여 weight 값을 조절해 줌으로써 각 단모음별로 학습시켰다. 이 때 sigmoid 함수에 사용된 desired output은 class에 속할 때 0.9이고 class에 속하지 않을 때 0.1을 주었다. 이에 대한 플로차트는 그림 1과 같다.

위에서 말한 것처럼 학습된 데이터는 입력 데이터가 들어 오면 각 단모음별로 MLP를 통과하여 그 결과값을 조사하여 가장 값이 큰 것으로 나타난 단모음이 인식된 것으로 간주하였다. 이에 대한 플로차트는 그림 2와 같다.

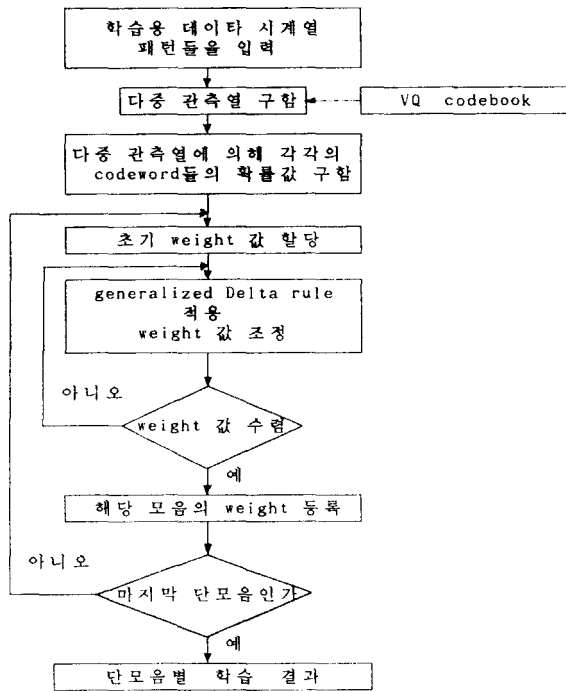


그림 1. Fuzzy를 이용한 VQ/NN의 학습 방법  
Fig 1. The learning method of VQ/NN using fuzzy.

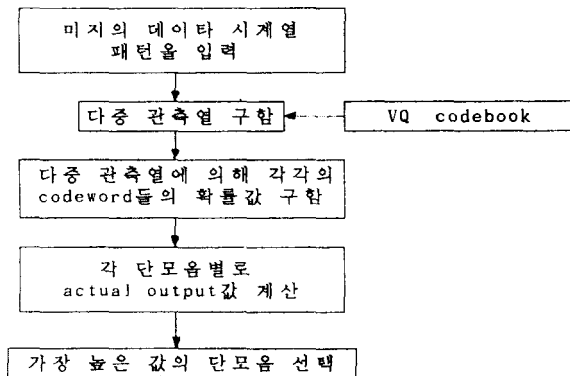


그림 2. Fuzzy를 이용한 VQ/NN의 인식 방법  
Fig 2. The recognition method of VQ/NN using fuzzy.

### III. 실험 및 결과

본 연구는 fuzzy를 이용한 VQ/NN에 의한 음성 인식 방법으로 실험에 사용된 데이터는 단모음 8자 /t/, /l/, /r/, /k/, /n/, /-/, /j/, /#/로써 10명의 화자가 10번 발음한 총 800개의 음성으로 이 중 3명이 각 단모음에 대해 5번씩 발음한 데이터를 가지고 학습을 시켰으며, 나머지 7명의 데이터를 가지고 실험하였다. 실험에서 같은 조건하에서 학습시킨 이유는 본 연구에서 제안한 방법이 얼마나 학습 능력이 좋은가를 실험해 보기 위해서이다. 따라서 본 연구에서 제안하는 방법이 외에도 비교를 위해서 일반적인 VQ/NN과 fuzzy를 이용한 VQ/HMM에 의한 방법도 실험하였다.

#### 1. 인식 시스템 구성

본 논문에서 제안하는 fuzzy를 이용한 VQ/NN에 의한 음성 인식 시스템은 그림 3와 같다. 데이터는 연구실에서 마이크를 사용하여 IBM-586에서 받았는데 3.5KHz의 LPF를 통과시킨 후 8KHz의 샘플링 주파수하에서 A/D 변환기를 통해 12비트로 양자화된 데이터이다. 이렇게 받아들인 데이터는 시작점과 끝점을 검출한 후 특징 벡터로 LPC cepstrum 계수를 구하고 이것을 VQ codebook의 codeword와 거리를 비교하여 다중관측열을 구한 후, 각 codeword가 가질 수 있는 확률을 계산해 이 codeword가 가질 수 있는 확률을 가지고 NN의 입력으로 삼아 학습 및 인식을 시켰다.

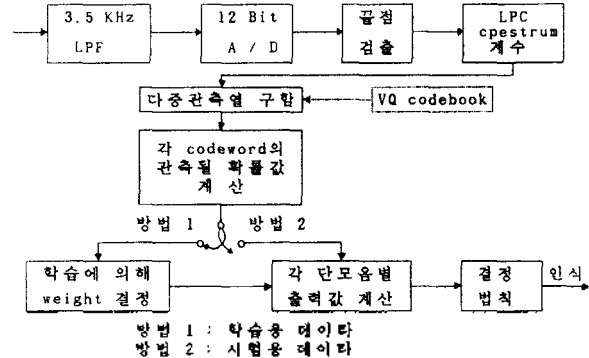


그림 3. Fuzzy를 이용한 VQ/NN 인식 시스템  
Fig 3. Recognition system of VQ/NN using fuzzy

#### 2. 인식 결과

##### 2.1 Fuzzy를 이용한 VQ/HMM에 의한 인식 결과

실험에서는 A, B, C 3명이 각 단모음에 대해 5번씩 발음한 것으로 학습시켰으며 인식 실험은 나머지 7명이 10번씩 발음한 8개 단모음 560개이다. 이 때 state수는 단모음 인식인 관계로 3으로 하였으며, 관측 심볼수(codebook size)는 32이다. 여기서 fuzzy rule은 시험 데이터의 프레

임당 다중 심볼수를 1에서부터 5까지 적용시켰다. 인식 실험에서 나타난 에러의 갯수를 표 1에 나타내었다. 실험 결과를 살펴 보면, 프레임당 심볼수를 하나만 선택(일반적인 HMM 인식)사에는 에러의 갯수가 67개이고, fuzzy

표 1. Fuzzy를 이용한 VQ/HMM에 의한 에러의 갯수  
Table 1. The number of errors by VQ/HMM using fuzzy

(a) 단일 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	1	3	0	1	4	6	0	0	15
E 화자	0	0	1	0	2	3	1	0	7
F 화자	0	0	7	0	2	1	1	0	11
G 화자	0	1	2	1	1	3	0	0	8
H 화자	0	4	0	2	3	1	0	0	10
I 화자	0	0	2	0	0	1	1	0	4
J 화자	2	0	4	0	3	1	2	0	12
전체	3	8	16	4	15	16	5	0	67

(b) 2개의 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	1	2	0	0	3	6	0	0	12
E 화자	0	0	1	0	1	3	0	0	5
F 화자	0	0	3	0	2	1	0	0	6
G 화자	0	0	2	1	0	3	0	0	6
H 화자	0	1	0	2	3	1	0	0	7
I 화자	0	0	2	0	0	1	1	0	4
J 화자	0	0	4	0	3	1	2	0	10
전체	1	3	12	3	12	16	3	0	50

(c) 3개의 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	1	2	0	0	3	4	0	0	10
E 화자	0	0	1	0	1	2	0	0	4
F 화자	0	0	1	0	2	1	0	0	4
G 화자	0	0	2	1	0	3	0	0	6
H 화자	0	1	0	2	2	1	0	0	6
I 화자	0	0	2	0	0	1	1	0	4
J 화자	0	0	1	0	3	1	1	0	6
전체	1	3	7	3	11	13	2	0	40

(d) 4개의 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	1	2	0	0	4	1	0	0	8
E 화자	0	0	1	0	1	2	0	0	4
F 화자	0	0	1	0	2	1	0	0	4
G 화자	0	0	0	1	0	3	0	0	4
H 화자	0	1	0	2	2	1	0	0	6
I 화자	0	0	2	0	0	1	1	0	4
J 화자	0	0	1	0	2	1	1	0	5
전체	1	3	5	3	11	10	2	0	35

(e) 5개의 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	1	2	0	0	5	1	0	0	9
E 화자	0	0	1	0	1	2	0	0	4
F 화자	0	0	1	0	2	1	0	0	4
G 화자	0	0	0	1	0	4	0	0	5
H 화자	0	1	0	2	2	1	0	0	6
I 화자	0	0	2	0	0	1	1	0	4
J 화자	0	0	1	0	2	1	1	0	5
전체	1	3	5	3	12	11	2	0	37

rule에 의해 다중 심볼수를 2개 선택시에는 50개, 3개 선택시에는 40개, 4개 선택시에는 35개, 5개 선택시에는 37개의 에러가 발생한다. 따라서, 다중 심볼수 4개일 때가 93.8%로 인식률이 가장 좋다.

3.2 제안된 fuzzy를 이용한 VQ/NN에 의한 인식 결과

본 연구에서 제안하는 실험에서도 HMM에서와 마찬가지로 A, B, C 3명이 각 단모음에 대해 5번씩 발음한 것으로 학습시켰으며, 인식 실험은 나머지 7명이 10번씩 발음한 8개 단모음 560개이다. 여기서 fuzzy rule은 시험 때 이타의 프레임당 다중 심볼수를 1에서부터 5까지 적용시켰다. 이 때 입력 노드수는 각 단모음별로 관측 심볼수와 마찬가지로 32이며 이들 노드의 값은 다중 관측 심볼들이 관측될 확률값이다. 인식 실험에서 나타난 에러의 갯수를 표 2에 나타내었다. 실험 결과를 살펴 보면, 프레임당 심볼수를 하나만 선택(VQ/NN 인식)사에는 에러의 갯수가 43개이고, fuzzy rule에 의해 다중 심볼수를 2개를 선택할 시에는 32개, 3개 선택시에는 26개, 그리고 4개 선택시에는 24개, 5개 선택시 25개의 에러가 발생한다. 따라서, 다중 심볼수 4개일 때가 95.7%로 인식률이 가장 좋다.

표 2. Fuzzy를 이용한 VQ/NN에 의한 인식률  
Table 2. The number of errors by VQ/NN using fuzzy

(a) 단일 심볼 선택시 (단위: 갯수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	0	1	0	0	2	3	0	0	6
E 화자	0	0	0	0	2	3	1	0	6
F 화자	0	0	3	0	1	1	0	0	5
G 화자	0	0	2	0	0	2	0	0	4
H 화자	0	2	0	2	3	1	1	0	9
I 화자	0	0	2	0	0	2	1	0	5
J 화자	1	0	5	0	2	0	0	0	8
전체	1	3	12	2	10	12	3	0	43

(b) 2개의 심볼 선택시 (단위: 개수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	0	1	0	0	2	3	0	0	6
E 화자	0	0	0	0	2	1	1	0	4
F 화자	0	0	1	0	0	1	0	0	2
G 화자	0	0	1	0	0	2	0	0	3
H 화자	0	0	0	2	3	1	0	0	6
I 화자	0	0	2	0	0	2	1	0	5
J 화자	1	0	2	0	3	0	0	0	6
전체	1	1	6	2	10	10	2	0	32

(c) 3개의 심볼 선택시 (단위: 개수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	0	1	0	0	1	3	0	0	5
E 화자	0	0	0	0	2	1	0	0	3
F 화자	0	0	0	0	0	1	0	0	1
G 화자	0	0	1	0	0	2	0	0	3
H 화자	0	0	0	0	3	1	0	0	4
I 화자	0	0	2	0	0	2	1	0	5
J 화자	0	0	2	0	3	0	0	0	5
전체	0	1	5	0	9	10	1	0	26

(d) 4개의 심볼 선택시 (단위: 개수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	0	0	0	0	1	3	0	0	4
E 화자	0	0	0	0	2	1	0	0	3
F 화자	0	0	0	0	0	1	0	0	1
G 화자	0	0	1	0	0	2	0	0	3
H 화자	0	0	0	0	2	3	0	0	5
I 화자	0	0	1	0	0	2	1	0	4
J 화자	0	0	2	0	2	0	0	0	4
전체	0	0	4	0	7	12	1	0	24

(e) 5개의 심볼 선택시 (단위: 개수)

단모음 화자	ㅏ	ㅣ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	전체
D 화자	0	0	0	0	1	3	0	0	4
E 화자	0	0	0	0	2	1	0	0	3
F 화자	0	0	0	0	0	1	0	0	1
G 화자	0	0	1	0	0	2	0	0	3
H 화자	0	0	0	0	3	3	0	0	6
I 화자	0	0	0	0	2	1	1	0	3
J 화자	0	0	2	0	2	1	0	0	5
전체	0	0	3	0	8	13	1	0	25

IV. 결 론

본 연구는 fuzzy를 이용한 VQ/NN에 의한 음성 인식에 관한 논문으로 VQ codebook으로 부터 다중 관측열을 구하고 이를 토대로하여 각 codeword가 데이터로부터 가질 수 있는 확률값을 계산하여 그 값을 입력값으로하여 MLP에 의해 인식하는 시스템을 제안하였다.

본 연구에서 제안된 방법은 fuzzy를 이용한 VQ/HMM에 의한 실험과 비슷하게 VQ codebook을 이용하여 다중 Codeword를 구한다는 점이다.

또한 본 연구의 목적은 fuzzy의 성능을 알아 보고 또한 HMM과 NN의 학습 능력을 비교 실험해 본 것으로 같은 조건하에서 얼마나 정확하게 분류해 내는지를 실험해 보는 것으로, 학습된 데이터에서는 모두 100% 인식률을 나타내었고 또한 같은 화자의 학습되지 않은 데이터에서는 99.2%의 인식률을 나타내었다. 그러나, 학습에 참여하지 않은 불특정 화자의 인식에 있어서는 차이를 나타내어, 두 방법에서 가장 인식률이 좋은 것으로 나타난 심볼수가 4개인 경우에 fuzzy를 이용한 VQ/HMM의 경우는 93.8%의 인식률을 나타내고, 일반적인 VQ/NN의 경우는 92.3%의 인식률을 나타내며, fuzzy를 이용한 VQ/NN의 경우는 95.7%의 인식률을 나타낸다.

따라서, 실험 결과에 의하면, 본 연구에서 제안하는 fuzzy를 이용한 VQ/NN에 의한 인식 실험이 fuzzy를 이용한 VQ/HMM이나 일반적인 VQ/NN에 의한 음성 인식보다 학습 능력에서 우수함을 입증하였다.

참 고 문 헌

1. 김순협, "한국어 음성의 분석과 자동 인식에 관한 연구," 박사 논문, 연세대학교 대학원, 1982. 12.
2. Y. Linde, A. Buzo, and R. M. Gray "An algorithm of Vector Quantizer Design", IEEE Trans. Comman, Vol. COM-28 pp. 84-95, Jan 1980.
3. L. R. Rabiner, B. H. Juang, "An Intorduction to Hidden Markov Models", IEEE ASSP MAGAZINE JAN. 1986.
4. L. R. Rabiner, S. E. Levinson, M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition", Bell System Technical Journal, Vol. 62 No. 4, April 1983.
5. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang. "Phoneme Recognition using Time-Delay Neural Network," IEEE Tran. of Acoustics, Speech and Signal Processing, Vol. 1, ASSP-37, March 1989.
6. T. Kohonen, G. Barna and R. Chrisley, "Statistical Pattern Recognition with Neural networks: Benchmarking Studies," IEEE, Proc. of ICNN, Vol. 1, pp. 61-68, July 1988.
7. C. E. Shannon, "A mathematical Theory of Communication", Bell Sys. Tech. J. 27, pp. 379-423, 623-656. 1948.
8. J. T. Tou, R. C. Gonzalez, Pattern recognition Principles, Addison-Wesley Publishing Company, Inc, 1974.
9. R. P. Lippmann, "An Introduction to Computing with Neural Nets," IEEE ASSP Magazine, 4-22, April 1987.
10. Yuh-Han Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley Publishing Company, 1989.
11. 안태욱, 김순협, "퍼지에 기초를 둔 HMM을 이용한 음성 인식", 대한전자공학회지, 제 28권 B편 제 12호, pp. 68-74. 1991. 6.

12. 안태옥 외2인, "VQ와 Multi-layer perceptron을 이용한 단  
모음 인식에 관한 연구", 한국음향학회지, 제 12권 제 1호,  
pp. 55-60, 1993. 2.

▲안 태 옥(Tae Ock Ann)  
1994년도 새13권 제 2E호 참조