

퍼지에 기초를 둔 HMM을 이용한 음성 인식

(Speech Recognition Using HMM Based on Fuzzy)

安 泰 玉*, 金 淳 協*

(Tae Ock Ann and Soon Hyob Kim)

要 約

본 논문은 화자 독립의 음성 인식에 관한 연구로써, 퍼지 (fuzzy)를 이용한 HMM(hidden markov model) 모델을 제안한다.

제안된 인식 방법에서는 VQ 코드북(codebook)으로부터 거리값이 작은 순으로 퍼지 법칙을 적용함으로써 적당한 확률값을 준 다중 관측열(multi-observation sequences)을 구한다. 그런 다음, 이 다중 관측열을 이용하여 HMM 모델을 작성하고, 인식시에는 관측 확률값이 가장 높은 것을 인식된 것으로 선택한다. 실험에 사용된 인식 대상어는 전국 146개 DDD 지역명이며, 특징 파라메타로는 10차 LPC cepstrum 계수를 사용하였다.

본 연구에서는 제안된 모델에 의한 인식 실험을 수행하는 것 이외에도 비교를 위하여 DP, MSVQ 및 일반적인 HMM에 의한 음성 인식 실험을 같은 조건하에서 같은 데이터로 수행하였다. 실험결과, 본 연구에서 제안한 퍼지를 이용한 HMM이 기존의 DP 방법이나 MSVQ 방법 및 일반적인 HMM 모델보다 인식률 및 인식 시간에서 우수함을 입증하였다.

Abstract

This paper proposes a HMM model based on fuzzy, as a method on the speech recognition of speaker-independent. In this recognition method, multi-observation sequences which give proper probabilities by fuzzy rule according to order of short distance from VQ codebook are obtained. Thereafter, the HMM model using this multi-observation sequences is generated, and in case of recognition, a word that has the most highest probability is selected as a recognized word. The vocabularies for recognition experiment are 146 DDD are names, and the feature parameter is 10th LPC cepstrum coefficients. Besides the speech recognition experiments of proposed model, for comparison with it, we perform the experiments by DP, MSVQ and general HMM under same condition and data.

Through the experiment results, it is proved that HMM model using fuzzy proposed in this paper is superior to DP method, MSVQ and general HMM model in recognition rate and computational time.

*正會員, 光云大學校 電子計算機工學科
(Dept. of Computer Eng., Kwangwoon Univ.)
接受日字: 1991年 7月 16日

I. 서 론

(※ 본 연구는 1989년부터 1992년까지의 한국과학재단 “목적 기초연구” 지원에 의한 것임.) 본 논문은 전국 146개 DDD 지역명을 인식 대상으로 한 화자 독립의 음성 인식에 관한 연구로 단어

단위의 인식 실험을 행한다. 이와 같이 인식 단위로 단어 단위의 음성 인식을 수행하는데 있어서, 기존의 DP에 의한 인식¹⁾은 인식 시간이 너무 길고 인식률도 높지 않으며, VQ²⁾나 MSVQ³⁻⁴⁾는 인식 시간 및 기억 용량은 적으나, 인식률이 떨어진다. 따라서, 다른 방법보다 인식률 및 계산 시간 등이 우수한 모델로 HMM 모델을 선정하여 실험한다.

그러나 일반적인 HMM⁵⁻⁸⁾의 경우는 모델을 작성하는데 많은 화자의 많은 학습용 데이터가 필요하며, 모델 학습시에 참여하지 않은 발음은 인식률이 현저하게 떨어질 수 있는데 본 연구에서는 관측열 (observation sequence)을 구할 시에 퍼지 개념을 도입한 다중 관측열 (multi-observation sequence)를 이용하여 HMM 학습에 이용함으로써 이런 단점을 보완한다.

제안된 HMM 모델의 다중 관측열을 구하는 방법은 거리값이 작은 것 몇개를 선택해서 거리값이 작은 순으로 퍼지 개념을 도입하여 적당한 가중치를 주는 방법이다. 따라서, 기존의 HMM에서는 학습시에 가장 거리값이 작은 것으로 선택되지 않아 확률값을 가지지 못하는 것이, 퍼지를 이용하는 경우에는 다중 관측열에 속하게 되면 어느 정도 확률값을 가질 수 있으므로, 인식시에 이것이 선택되며 확률값을 가지므로 인식률을 향상시킬 수 있다. 물론 기존의 방법에서도 smoothing 기법을 이용하여 관측 심볼로 선택되지 않아도 확률이 0가 되는 것을 방지하고 있다. 그러나 그것만으로는 미흡한 방편에 지나지 않아 인식에 많은 도움을 주지 못한다.

본 연구에서는 ZCR(zero crossing rate)과 에너지(energy)로 끝점 검출을 행했으며, 특징 파라미터로는 LPC 켈스트럼 계수를 이용하였다.

II. 제안된 음성 인식 시스템

일반적인 이산 HMM의 경우는 각 열마다 하나의 심볼만 관측되는 것으로 생각하였는데 본 연구에서는 각 열마다 몇개의 다중 심볼이 퍼지 개념에 의해 적당한 가중치를 가지고 한 열을 구성하는 다중 관측열을 사용하였다. 그 이유는 HMM 학습시에 선택되지 않은 심볼은 학습 과정에서 확률값을 가질 수 없는 관계로 인식 실험시 학습에 참여하지 않는 심볼이 나올 경우에는 인식이 거의 불가능하기 때문이다. 따라서, 다중의 관측열을 둬으로써 같은 열에서 몇개의 관측될 다중 심볼에 속하면 모델링시에 확률값을 가질 수 있으므로 인식시에 어느 정도의 확률값을 가져 인식될 가능성을 높여준다.

1. 퍼지 개념¹⁾을 이용한 다중 관측열 나열

각 프레임의 벡터와 VQ 코드북의 코드워드 중 거리가 가장 작은 것을 심볼로 선택하는 것이 일반적인 HMM에서 관측열을 구하는 방법이었다. 그런데, 본 연구에서는 일반적인 HMM에서 관측열을 구하는 것과는 달리 각 열마다 몇개의 유사한 특징을 가지는 심볼을 선택하는 방법을 제안한다.

실험에서 사용된 다중 관측열을 구하는데 사용한 1형 퍼지 법칙은 다음과 같다.

$$w_s = \frac{2^S - s}{\sum_{s=1}^S 2^{S-1}} \quad (1)$$

여기서 s는 어떤 프레임의 벡터와 코드북의 각 코드워드 중 거리가 작은 순으로 표시했을 때 몇 번째인지를 나타내고, S는 한 열에서 선택될 다중 심볼수를 의미하며, w_s는 전체 확률을 1로 했을 때 거리가 s번째인 심볼이 가질 가중치이다.

본 연구에서는 다중 심볼수 S의 값을 변경하며 실험한다. 1형 퍼지에 의한 실험을 행한 후 가장 인식률이 좋은 다중 심볼수 값이 나오면 제안된 2형 퍼지에서는 이것을 바탕으로 실험을 행한다.

1형 퍼지에서는 거리값이 작은 순으로 무조건 식(1)처럼 확률값을 주었는데 실제에 있어서는 거리값에 따르는 고려가 있어야 한다는 개념에서 2형 퍼지를 사용하였다. 따라서, 2형 퍼지에서는 식(1)에서 구한 w_s의 값들을 다음 식(2)와 같이 변경시킨다.

$$\bar{w}_s = \begin{cases} w_s + (S_0 * (s-1) - (d_s/d_1 - 1)) / S_0 * w_s * S_w & (1 + S_0 * (s-2) < d_s/d_1 < 1 + S_0 * s) \\ w_s * (1 + S_w) & (1 + S_0 * (s-2)) > d_s/d_1 \\ w_s * (1 - S_w) & d_s/d_1 > 1 + S_0 * s \end{cases} \quad (2)$$

여기서, d₁은 첫번째로 거리값이 작은 심볼의 거리값을 의미하고, d_s는 s번째 심볼의 거리값을 의미한다. 또한, 위의 수식에서 S₀는 첫번째 선택될 심볼과 두번째 선택될 심볼간에 거리값의 비율이 1+S₀이며 두번째 선택될 심볼이 1형 퍼지에서 취한 확률값과 같은 값을 가질 수 있도록 해주며, 만약 거리값의 비율이 1+S₀보다 더 크면 두번째 선택될 심볼의 가중치가 작아지고, 반대로 비율이 1+S₀보다 작아지면 두번째 심볼의 가중치가 커지게 되는 파라메타이고, 마찬가지로 S₀는 세번째 선택될 심볼은 첫번째 심볼과의 거리값이 비가 1+2*S₀이며 1형 퍼지에서 가지는 확률값과 같은 값을 가질 수 있도록 해주며, 또한 가중치가 커지고 작아지는 것은 앞의 경우와 마

찬가지의 원리에 의해 이루어지는데, 나머지 선택될 심볼들도 마찬가지로 법칙에 의해 값이 결정된다. 그리고, S_w 는 가중치가 변할 수 있는 비율의 폭을 의미한다. 이 식(2)에 의해 구해진 값은 전체의 값이 1이 되도록 다음과 같이 식(3)에 의해 조정된다.

$$w_s = \frac{\bar{w}_s}{\sum_{i=1}^S w_i} \quad (3)$$

2. 제안된 방법에 의한 음성 인식

본 연구에서는 각 출력 심볼에 퍼지 개념을 도입하여 가중치를 줌으로서 인식률을 향상시키는 방법을 제안한다.

1) 2형 퍼지를 이용한 HMM의 원리

본 연구에서 제안하고 있는 2형 퍼지 개념을 도입한 음성 인식 방법에서 사용되는 기호는 다음과 같다.

상태수 : N

전체 심볼수 : M

열의 관측될 심볼수 : S

상태 집합 : $Q = \{q_1, q_2, \dots, q_N\}$

심볼 집합 : $V = \{v_1, v_2, \dots, v_M\}$

관측열의 길이 : $t = 1, 2, \dots, T$

t번째 관측 심볼열이 상태 q_t 에 있고, t+1번째 관측 심볼열이 상태 q_t 를 선택할 확률

$$A = \{a_{ij}\}, a_{ij} = pr(q_j \text{ at } t+1 | q_i \text{ at } t) \quad (4)$$

$(1 \leq i, j \leq N)$

t번째 관측 심볼열이 q_j 상태에서 다중 심볼 집합 $\{v_{k1}, v_{k2}, \dots, v_{ks}, v_{ks}\}$ 을 가지고 그때의 각각의 심볼이 가질 가중치 집합을 $\{w_{k1}, w_{k2}, \dots, w_{ks}, \dots, w_{ks}\}$ 이라 하면

$$\sum_{s=1}^S w_s(k_s) = 1 \quad (5)$$

이다. 따라서, t번째 관측 심볼열이 q_j 상태에서 다중 심볼 $v_k = \{v_{ks}\}$ 를 선택할 확률

$$\begin{aligned} B &= \{b_j(k) | b_j(k) = \sum_{s=1}^S w_s(k_s) b_{js}(k_s) \\ &= pr(v_k \text{ at } t | q_j \text{ at } t) \\ &= pr(\{v_{ks}\} \text{ at } t | q_j \text{ at } t) \\ &(1 \leq j \leq N), (1 \leq k \leq M), (1 \leq s \leq S) \end{aligned} \quad (6)$$

초기 상태에서 상태 q_1 를 선택할 확률

$$\pi = \{\pi_i\}, \pi_i = pr(q_i \text{ at } t=1) \quad (7)$$

관측열 $O = \{O_1, O_2, \dots, O_T\}$

관측열 심볼의 구성 $O_t = \{o_{t1}, o_{t2}, \dots, o_{ts}\}$

이상의 정의를 이용한 모델은 $\lambda = (A, B, w, \pi)$ 로 표시할 수 있는데 이 모델을 실제 응용하는 데는 모델 작성시에 사용되는 알고리즘과 인식시에 사용되는 알고리즘으로 나누어 생각할 수 있다.

(1) 모델 학습 알고리즘

능률적인 계산방법인 L. E. Baum에 의해 제안된 전향-후향 알고리즘^[5]을 수정한 알고리즘을 간단히 살펴보자. 먼저 전향 변수 $\alpha_t(i)$ 을 살펴보면

$$\begin{aligned} \alpha_t(i) &= Pr(O_1, O_2, \dots, O_t = q_i | \lambda) \\ (O_t &= \{o_{t1}, o_{t2}, \dots, o_{ts}\}) \end{aligned} \quad (8)$$

로 정의 한다.

이는 주어진 모델 λ 에 대해서 시간 t에서 관측열이 O_1, O_2, \dots, O_t 이고 상태가 q_i 일 확률이며 다음과 같은 절차에 의해서 구할 수 있다.

단계 1. 초기화

$$\begin{aligned} \alpha_1(i) &= \pi_i b_i(O_1) = \sum_{s=1}^S \pi_i w_s(O_{1s}) b_{is}(O_{1s}) \\ 1 \leq i \leq N \end{aligned} \quad (9)$$

단계 2. $t=1, 2, \dots, T-1$ 에 대해 반복 ($1 \leq i \leq N$) N

$$\alpha_{t+1}(j) = \sum_{s=1}^S \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] w_s(O_{t+1s}) b_{js}(O_{t+1s}) \quad (10)$$

단계 3. 그러면

$$P(O | \lambda) = \sum_{i=1}^N \alpha_{T0} \quad (11)$$

같은 방법으로 후향 변수를 살펴보면,

$$\begin{aligned} \beta_t(i) &= Pr(O_{t+1}, O_{t+2}, \dots, O_{T-1}, O_T | i_t = q_i, \lambda) \\ (O_t &= \{o_{t1}, o_{t2}, \dots, o_{ts}\}) \end{aligned} \quad (12)$$

로 정의할 경우 다음 절차에 의해 구할 수 있다.

단계 1. 초기화

$$\beta_T(i) = 1, 1 \leq i \leq N \quad (13)$$

단계 2. $t=T-1, T-2, \dots, 1$ 에 대해 반복 ($1 \leq i \leq N$)

$$\begin{aligned} \beta_t(i) &= \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \\ &= \sum_{j=1}^N \sum_{s=1}^S a_{ij} w_s(O_{t+1s}) b_{js}(O_{t+1s}) \beta_{t+1}(j) \end{aligned} \quad (14)$$

이다. 일반적인 HMM에서 초기 파라메타들로부터 $Pr(O | \lambda)$ 를 최대로 하는 $\lambda = (A, B, w, \pi)$ 를 재추정하는 것으로 Baum-Welch 재추정 알고리즘^[5]을 이용하는데, Baum-Welch의 재추정 알고리즘도 전향-후향 알고리즘과 마찬가지로 일반적인 알고리즘에서 수정하여 사용할 수 있다. 주어진 관측열과 모델 λ 에 대해서 시간 t에서의 상태가 q_i 이고 시간 t+1에서의

상태가 q_t 일 확률 $\xi_t(i, j)$ 를 다음과 같이 정의한다.

$$\xi_t(i, j) = Pr(i_t = q_i, i_{t+1} = q_j | O, \lambda) \quad (15)$$

$$= \frac{\sum_{s=1}^S \alpha_t(i) a_{i,j} w_s(O_{t+1}) b_{j,s}(O_{t+1}) \beta_{t+1}(j)}{Pr(O | \lambda)}$$

또한 주어진 관측열과 모델 λ 에 대해서 시간 t 에서 상태가 q_t 일 확률은 다음과 같다.

$$\gamma_t(i) = Pr(i_t = q_i | O, \lambda) = \sum_{j=1}^{T-1} \xi_t(i, j) \quad (16)$$

여기서, 현재 모델 $\lambda = (A, B, w, \pi)$ 이라 하고, 다시 추정된 모델은 $\hat{\lambda} = (\hat{A}, \hat{B}, w, \pi)$ 이라 하면 새로운 모델 파라메타는 아래의 식들과 같다.

$$\hat{\pi}_1 = \gamma_1(i), 1 \leq i \leq N \quad (17)$$

$$\hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (18)$$

$$= \frac{\sum_{t=1}^{T-1} \sum_{s=1}^S \alpha_t(i) a_{i,j} w_s(O_{t+1}) b_{j,s}(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i)}$$

$$\hat{b}_j(k) = \frac{\sum_{t=1, O_t=O_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} = \frac{\sum_{t=1, O_t=O_k}^T \alpha_t(j) \beta_t(j)}{\sum_{t=1}^T \alpha_t(j) \beta_t(j)} \quad (19)$$

2) 인식 알고리즘

인식에 사용되는 알고리즘으로는 전향 알고리즘과 후향 알고리즘 외에 Viterbi 알고리즘^[5]이 있다. 본 연구에서는 논문(7)의 실험 결과 전향 알고리즘이 인식률이 더 좋았으므로 전향 알고리즘을 이용하였다.

3) HMM 모델 작성 및 인식 시스템의 구조

제안된 방법은 학습용 데이터들의 다중 관측열을 이용하여 HMM 모델의 파라메타들을 학습 시킨다. 이에 대한 순서도는 그림1에 나타내었다. 또한 인식 시에는 두가지 방법으로 인식하는데 첫번째 방법은 모델 작성시와 마찬가지로 다중 관측열에 의해 인식하는 방법이고, 두번째 방법은 모델 작성시와는 달리 일반적인 HMM에서 처럼 그냥 단일 관측열을 구해 인식하는 방법이다. 이에 대한 순서도는 그림2에 나타내었다.

본 모델에서도 일반적인 HMM과 마찬가지로 세가지 문제점이 대두된다. 첫째는 모델의 파라메타를 어떻게 초기화 할 것인가 하는 점인데, 이것은 처음에 각 파라메타의 값을 균등한 값에다 아주 작은 값의 변화를 줌으로써 초기화하여 해결한다. 둘째는 관측열의 길이 T 가 증가함에 따라 이들의 값이 지

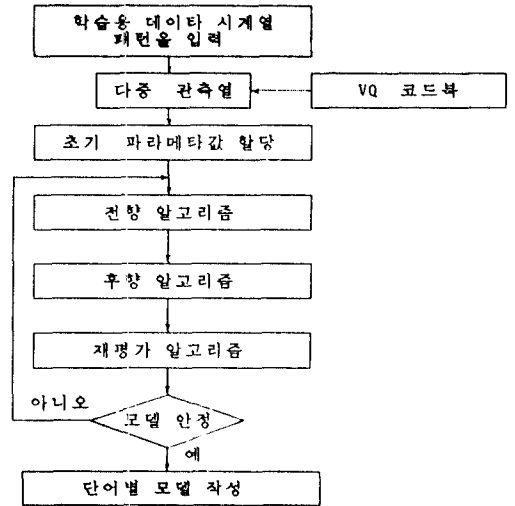


그림 1. 제안된 시스템의 모델 작성법

Fig. 1. Method of model generation of proposed system.

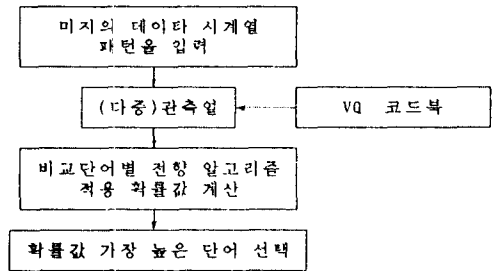


그림 2. 제안된 시스템의 인식 방법

Fig. 2. Recognition method of proposed system.

수함수적으로 감소하여 곧 언더플로우가 발생하게 된다는 점인데, 이것은 scaling 기법에 의해 해결한다. 세째는 학습 데이터의 양이 충분히 많지 않을 경우 학습 과정에서 다중 관측열에 의해 학습 시키더라도 어떤 심볼이 모델 작성시에 나타나지 않을 것으로 추정되었는데 인식시에 이 심볼이 나타나는 경우가 있을 수 있다는 점인데, 이것은 smoothing 기법을 이용하여 학습시에 나타나지 않은 심볼에도 아주 작은 확률값을 줌으로써 해결한다. 또한, 본 연구에서는 천이 가능수가 1인 left-to-right 모델을 사용하였다.

III. 인식 실험 결과

본 연구는 인식 실험 대상어휘로 대한민국 전역

146개 DDD 지역명을 선정하였고, 5명의 남성 화자에 의해 3번씩 발성된 것 중에서 각각 2번씩 발음한 것으로 모델을 작성하였으며 나머지 각각 1 번씩 발음한 730개의 발음으로 인식 실험을 수행하고 다른 인식 방법들과 비교하였다.

1. 인식 시스템 구성

본 연구에서는 실험한 음성 인식 시스템은 그림 3과 같다. 특징 벡터는 LPC 계수보다 LPC cepstrum 계수가 더 좋다는 고찰⁹⁻¹⁰⁾에 따라 LPC cepstrum 계수를 사용하였다.

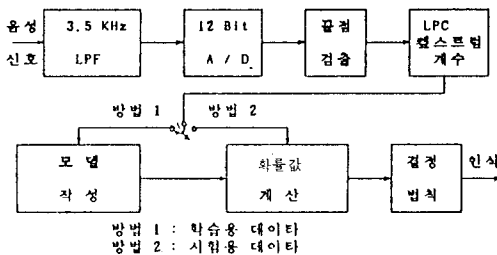


그림 3. 인식 시스템의 구성도

Fig. 3. Block diagram of recognition system.

2. 기존의 인식 방법에 의한 실험 결과

DP 패턴 매칭의 경우, 표준 패턴을 선택하는 방법은 집단화 알고리즘을 사용하였고, 소구간 경로 제약 및 전체 경로 제약은 Sakoe와 Chiba의 방법¹¹⁾을 사용하였으며, MSVQ에 의한 음성 인식의 경우¹²⁾는 구간 수를 8로 하여 실험하였으며, 각 구간별 코드북의 크기는 4로 하여 총 32개의 코드워드를 택하여 실험하였다.

또한, 일반적인 HMM의 경우는 256개의 코드워드(심볼)를 주었으며, 상태수는 8로 하였다.

이 세가지 방법에 의한 인식 실험 결과는 표 1과 같다.

표 1. 기존의 방법에 의한 실험 결과

Table 1. Results by classical methods.

방법 인식	DP 인식 방법	MSVQ 인식 방법	HMM 인식 방법
인식률	77.40	72.74	79.45

3. 제안된 인식 방법에 의한 실험 결과

본 연구에서는 제안한 방법에 의한 인식 실험에서는 state 8과 codeword 수는 256으로 실험하였다.

1) 1형 퍼지를 이용한 HMM 실험 결과

가중치를 가질 관측 심볼수 S는 2에서 8까지 선택하여 학습시켰고, 인식시에는 두가지 방법으로 실험하였다.

첫번째 방법은 학습시와 마찬가지로 인식시에도 다중 관측 심볼을 선택하여 실험하였고, 두번째 방법은 인식시에 각 열마다 관측 심볼수를 1개만 선택하여 실험하였는데 그 결과는 표2에 나타내었다.

표 2. 1형 퍼지를 이용한 HMM의 인식 결과

Table 2. Result of HMM using 1 type fuzzy.

인식 방법	관측될 심볼 수						
	2	2	4	5	6	7	8
다중관측심볼	87.54	89.05	88.91	88.91	89.73	89.59	88.90
단일관측심볼	79.878	87.40	88.22	89.45	90.25	90.00	90.00

2) 제안된 2형 퍼지를 이용한 HMM 실험 결과

1형 퍼지를 이용한 HMM 음성 인식에서 관측될 심볼수가 6인 경우가 가장 인식률이 좋으므로 심볼수가 6인 경우를 중심으로 식(2)와 식(3)을 이용하여 실험하였다.

표 3. S₀의 변화에 따른 HMM의 인식 결과

Table 3. Result of HMM by variation S₀.

인식 방법	S ₀ 의 값		
	0.25	0.5	0.75
다중관측심볼	89.59	89.45	89.59
단일관측심볼	90.30	90.41	90.14

표3은 학습시에 심볼수를 6으로 하고 S_w를 0.2로 하였으며 S₀를 0.25, 0.5, 0.75로 변화시키면서 실험하였다. 여기서 표2의 경우와 표3의 경우에서 볼 수 있는 바와 같이 학습시에 다중 심볼을 이용하여 학습시키고, 인식시에는 단일 심볼을 이용하여 인식하는 것이 좋다. 따라서, 이후의 실험에서는 인식시에 단일 심볼만을 이용하여 실험을 행한다.

다음은 다중 심볼을 6으로하고, S₀를 0.5로 고정시킨 상태에서 S_w의 값을 0.1, 0.2, 0.3, 0.5로 변화시키면서 실험하였는데, 표4에서 보는 바와 같이 S₀가 0.5일 때 가장 인식률이 좋다.

표 4. S_w 의 변화에 따른 HMM의 인식 결과

Table 4. Result of HMM by variation of S_w .

인식 방법	S_w 의 값			
	0.1	0.2	0.3	0.5
단일관측심볼	90.41	90.41	90.00	89.45

또한, S_o 를 0.5로 하고, S_w 를 0.2로 한 상태에서 다중 심볼수가 7과 10인 경우를 실험하여 그 결과를 표5로 나타내었는데, 2형 퍼지에서도 역시 다중 심볼수가 6일 때 가장 인식률이 좋았다.

표 5. 다중 심볼수에 의한 HMM 인식 결과

Table 5. Result of HMM by number of multi-symbol.

인식 방법	S_o 의 값			
	0.2	0.3	0.4	0.5
단일관측심볼	89.73	90.14	90.55	90.41

표4와 표5의 실험 결과에서 다중 심볼수가 6이고, S_o 가 0.5이며, S_w 가 0.1과 0.2인 경우가 인식률이 가장 좋을음을 알 수 있었다. 따라서, S_w 가 0.1인 경우에 S_o 를 0.2에서 0.5까지 변화시키면서 실험 하였다. 그 결과를 표6에 나타내었다.

표 6. $S_w=0.1$ 에서 S_o 의 변화에 따른 HMM 인식 결과

Table 6. Result of HMM by variation of S_o in $S_w=0.1$.

인식 방법	방법	다중 심볼수		
		6	7	10
단일관측심볼		90.40	90.27	89.73

지금까지의 실험 결과를 종합해 볼 때, 2형 퍼지를 이용한 인식에서 가장 인식률이 좋은 것으로 나타나는 factor들은 다중 심볼수가 6이고, S_o 가 0.4이며 S_w 가 0.1인 경우이다.

4. 종합적인 실험 결과

본 연구에서는 제안한 음성 인식 방법을 일반적인 HMM 및 다른 인식 방법과 비교하였다. 이 때 비교되는 것은 인식 방법에서 가장 좋은 인식률을 나타내는 것으로 기억 용량과 인식시 처리 속도등을 포함하여 표7에 나타내었다.

표 7. 종합적인 실험 결과

Table 7. All-round experimental result.

방법	구분	기억용량	처리 속도	인식률 (%)
MSVQ	51,392	515,088 multiply	72.74	
HMM	311,168	206,080 multiply 93,440 log	79.15	
1형 퍼지를 이용한 HMM	311,168	206,080 multiply 93,440 log	90.25	
세단된 2형 퍼지를 이용한 HMM	311,168	206,080 multiply 93,440 log	90.55	

여기에서 기억 용량 및 처리 속도는 Rabiner등^[8]에 의한 계산식을 사용하였다. 기억 용량 및 처리 속도 계산시에 평균 프레임 수는 40이다.

IV. 결 론

HMM 모델을 이용한 음성 인식을 하는데 있어서 모델의 파라메타를 잘 학습시키는 것은 중요한 일이다. 따라서, 본 연구에서는 퍼지를 이용한 HMM모델을 제안하고, 이 모델을 이용하여 화자 독립의 음성 인식 실험을 수행하였고 일반적으로 많이 사용되고 오고 있는 DP 방법과 MSVQ 방법에 의한 인식 실험 및 일반적인 HMM에 의한 인식 실험을 같은 조건하에서 실행하여 비교하였다.

기존의 대표적인 인식 방법인 DP 패턴 매칭 방법의 경우는 계산 시간도 많이 걸리고 인식률도 높지 않은 단점이 있고, MSVQ의 경우는 기억 용량이 적고 인식 시간이 적게 걸리나, 인식률이 낮은 단점이 있으며, 또한 일반적인 HMM의 경우는 모델을 만드는 데 시간이 많이 걸릴 뿐만 아니라 학습하는데 많은 데이터가 필요하다.

그런데 본 연구에서 처럼 각 프레임의 한 특징 벡터에 대해서 하나의 심볼만 선택하여 학습한 후 인식할 경우에 발생할 수 있는 오인식을 퍼지 개념의 다중 관측열을 이용하여 모델을 작성함으로써 줄여 주고 인식률을 향상시킨다.

오인식된 단어들을 살펴보면, 비슷한 단어들간의 애매한 발음(예를 들면, 대천을 대전에 가깝게, 성주를 상주에 가깝게, 양양을 영양에 가깝게 등)과 잡음이 심한 경우로 D/A에 의해 들어 보아도 구별이 가지 않는 경우가 많았다.

따라서, 화자가 정확하게 발음해야 함은 물론 잡음에 잘 적응하고 보다 더 나은 인식률을 얻을 수 있는 인식 시스템의 개발이 필요하고, 한국어의 음운학

적인, 그리고 음향학적인 특성에 맞는 알고리즘의 개발이 필요하다.

參 考 文 獻

- [1] Hiroaki Sakoe and Seibi Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-26, no. 1, pp. 43-49, Feb. 1978.
- [2] R.M. Gray, "Vector quantization," *IEEE ASSP Magazine*, Apr. 1984.
- [3] Tae Ock Ann and Sun hyub Kim, "An automatic speech recognition of computer using time sequential vector quantization," *The Korea Institute of Telematics and Electronics*, vol. 27, no. 7, July. 1990.
- [4] D.K. Burton, J.E. Shore and J.T. Buck, "Isolated-word speech recognition using multisection vector quantization codebooks," *IEEE Trans. on Acoustics, Speech, Signal Processing*, vol. ASSP-33, no. 4, Aug. 1985.
- [5] L.R. Rabiner and B.H. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, Jan. 1986.
- [6] Kai-Fu Lee, The Development of the

SPHINX System, Kluwer Academic Publishers, 1989.

- [7] T.O. Ann, Y.G. Byun and S.H. Kim, "Korean speech recognition using DHMM," *Acoustical Society of Korea*, vol. 10, no. 1, pp. 52-61, Feb. 1991.
- [8] L.R. Rabiner, S.E. Levinson and M.M. Sondhi, "On the application of vector quantization and hidden markov model to speaker-independent, isolated word recognition," *Bell System Technical Journal*, vol. 62, no. 4, Apr. 1983.
- [9] Shikano, K. and Kohda, M., "On the LPC Distance Measures for vowel recognition in continuous utterance," *Institute of Electrical and Communication Engineers of Japan, Trans. on D,J* 63-D, May. 1980.
- [10] K.Katsuyuki NIYADA, Satoru FUJII and Shuji MORII, "Evaluation of Feature Parameters and distance measures for speaker-independent speech recognition," *Institute of Electronic and Communication Engineers of Japan, Trans on A*, vol. J 69-A, no. 5, May. 1986.
- [11] 박민용, 최항식, 퍼지 시스템의 응용 입문, 대영사, 1990.

著 者 紹 介



安 泰 玉 (正會員)

1953年 6月 24日生. 1981年 2月
울산공과대학 재료공학과 졸업.
1987年 2月 광운대학교 대학원
전자계산기공학과 공학석사 학위.
현재 광운대학교 전자계산기공학과
박사과정 수료. 주관심분야는
다시탈 신호처리, 음성 인식, 인공 지능, 뉴로퍼지, 데
이타 통신 등임.



金 淳 協 (正會員)

1947年 12月 28日生. 1974年 2月
울산공과대학 전기공학과 졸업.
1976年 2月 연세대학교 전자공
학과 공학석사 학위 취득. 1983
年 2月 연세대학교 전자공학과
공학박사 학위 취득. 1986年 8
月~1987年 7月 The University of Texas at Austin
전기 및 전자공학과 객원 교수. 현재 광운대학교 공
과대학 전자계산기공학과 교수. 주관심분야는 디지
탈 신호처리, 음성 인식, 인공 지능 등임.