

DMS 모델과 퍼지 개념을 이용한 HMM에 기초를 둔 음성 인식

안태옥^{1*}

HMM-based Speech Recognition using DMS Model and Fuzzy Concept

Tae-Ock Ann^{1*}

요약 본 논문은 화자 독립의 음성인식을 위한 연구로서, DMS(Dynamic Multi-Section) 모델에 의한 DMSVQ(Dynamic Multi-Section Vector Quantization) 코드북과 퍼지 개념을 이용한 HMM(Hidden Markov Model) 음성인식 방법을 제안한다. 제안된 인식 방법에서는 학습 데이터를 동적으로 몇 개의 구간(section)으로 분할한 후, 각 구간마다 DMSVQ 코드북(codebook)으로 부터 거리값이 작은 순으로 퍼지 법칙을 적용함으로써 적당한 확률값을 준 다중 관측열(multi-observation sequences)을 구한다. 그런 다음, 이 다중 관측열을 이용하여 HMM을 작성하고, 인식시에는 관측 확률값이 가장 높은 것을 인식된 것으로 선택한다.

제안된 방법에 의한 인식 실험은 기존의 다양한 인식 실험들과 비교를 위해 동일한 조건하에서 같은 데이터로 수행하였다. 실험 결과로서, 본 연구에서 제안한 방법이 기존의 방법들보다 우수한 방법임을 입증하였다.

Abstract This paper proposes a HMM-based recognition method using DMSVQ(Dynamic Multi-Section Vector Quantization) codebook by DMS(Dynamic Multi-Section) model and fuzzy concept, as a study for speaker-independent speech recognition.

In this proposed recognition method, training data are divided into several dynamic section and multi-observation sequences which are given proper probabilities by fuzzy rule according to order of short distance from DMSVQ codebook per each section are obtained. Thereafter, the HMM 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.

Other experiments to compare with the results of recognition experiments using proposed method are implemented as a data by the various conventional recognition methods under the equivalent environment. Through the experiment results, it is proved that the proposed method in this study is superior to the conventional recognition methods.

Key Words : DMS, DMSVQ, fuzzy concept, multi-observation

1. 서론

본 논문은 음성 다이얼링(voice dialing) 시스템 개발을 위하여 대한민국의 기차역명 201개를 인식대상으로한 화자 독립의 한국어 단독어 인식에 관한 연구를 행하는 것이다. 단어 단위의 음성 인식을 하는데 있어서, 기존의 DP에 의한 인식^[1]은 인식 시간이 너무 길고 기억 용량도

크며, VQ^[2]나 MSVQ^[3-6]는 인식 시간 및 기억 용량은 적으나, 인식률이 떨어진다. 따라서 MSVQ 방법을 개선시킨 방법으로 구간을 동적으로 나누는 DMS 모델을 이용한 DMSVQ에 의한 인식 방법^[5]이 MSVQ에 의한 인식 방법보다 인식률이 높다.

그러나 1980년대 이후에는 DP나 VQ를 이용한 인식 방법보다 인식률이 우수한 HMM 모델을 이용한 인식 방

본 논문은 호원대학교 교내연구비로 수행되었음.

¹호원대학교 컴퓨터·계입학부

접수일 08년 6월 5일

수정일 08년 7월 21일

*교신저자 안태옥(youngandi@yahoo.co.kr)

계재확정일 08년 8월 11일

법^[6-7]이 지속적으로 인식 시스템에 적용되어왔다. HMM 모델을 이용한 인식 방법에 있어서도 시간 정보를 가지는 MSVQ 코드북을 이용한 인식 방법^[8]이 VQ 코드북을 이용한 인식 방법보다 인식률이 증가된다. 따라서 본 연구에서는 MSVQ보다 DMS 모델을 이용한 음성 인식 방법이 인식률이 더 높다는 점^[9]과 모델 학습시에 참여하지 않은 화자의 음성은 인식률이 현저하게 떨어질 수 있는데 관측열(observation sequence)을 구할 시에 퍼지 개념을 도입한 다중 관측열^[9]를 이용하며 이런 단점을 보완할 수 있다는 점에 착안하여 DMS 모델과 퍼지 개념을 이용한 HMM 음성 인식을 행하려 한다.

또한, 본 연구에서는 인식 실험을 행하는데 있어서 구간마다 확률값을 비교하여 인식 대상 단어의 후보수를 줄이고, 계속해서 거듭되는 구간마다 후보자 수를 줄이면서 이를 반복하여 최종적으로는 단어를 인식하는 DMS 모델과 퍼지 개념을 이용한 HMM 음성 인식을 제안한다.

2. 제안된 시스템

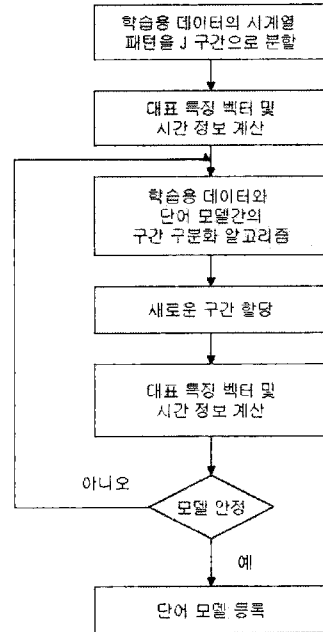
기존의 HMM이 코드워드의 벡터값과 미지단어의 어떤 프레임의 벡터간의 거리값이 가장 작은 것 하나를 택하는데 반해, 제안된 음성 인식 방법은 퍼지 개념을 이용한 HMM 모델의 다중 관측열을 구하는 방법으로 거리값이 짧은 몇개를 선택해서 거리값이 작은 순으로 적당한 가중치를 주고, 몇 개의 구간으로 구간을 동적으로 잡아 줌으로써 같은 특징의 정보가 같은 구간에 위치하게 함으로써 인식률을 높이는 DMS 모델과 퍼지 개념을 이용한 HMM 인식 방법이다.

2.1 DMS 모델 작성 방법

DMS 모델은 VQ 개념을 확장시킨 MSVQ 방법이 각 구간을 등간격으로 분할하여 프레임수를 동일한 길이로 잡아주는 것과는 달리 구간의 경계를 동적으로 잡아 각 구간의 프레임수를 가변 길이로 잡아 주는 모델이다.

모델 작성 방법은 구간 구분화 알고리즘^[5]의 결과로 등록된 각 학습용 데이터의 단어 패턴 (T_1, T_2, \dots, T_N) 들의 구간 경계를 이용하여 각 단어 패턴들을 J 개의 구간으로 나누어 주고, 단어 패턴의 j 번째 구간에 할당된 프레임들의 특징 벡터들을 모아 중심점을 계산해 이것을 단어 모델의 j 번째 구간을 대표하는 특징 벡터로 한다.

지속 시간 정보는 j 번째 구간의 마지막 프레임수들의 합을 단어 패턴들의 전체 프레임수로 나누어서 구한다. DMS 모델에 의한 단어 모델 작성 방법은 [그림 1]과 같다.



[그림 1] 단어 모델 작성 방법

2.2 퍼지 개념을 이용한 다중 관측열 나열

퍼지 이론의 개념을 음성 인식에 도입함으로써 화자의 차이와 시간적인 변동을 흡수 할 수 있다.

같은 음성에서도 발생 속도의 차이가 원인으로 음성의 길이가 일정하지 않다. 같은 사람이 같은 말을 하여도 그때마다 그 길이가 바뀐다. 물론 똑같은 말이라도 발생하는 사람이 다르며 길이의 변동 역시 크다. 또한 발생 기관의 크기는 인간에 따라 달라 같은 형태로 하여 발생하여도 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. 이들 중 일반적인 HMM의 경우는 많은 화자의 많은 학습 데이터를 통계적인 처리를 행함으로써 이를 극복하려 했다. 그런데, 이런 노력에도 불구하고 이것을 완전히 해결하지 못함으로 본 연구에서는 퍼지 이론을 이용하여 극복하려 한다. 어떤 발음을 했을 때 발음할 때마다 또는 화자에 따라 특성이 달라지지만 그 발음이 가지는 특성이 많이 벗어나지는 않을 거라는데 유의하여 본 연구에서는 퍼지 이론을 이용한다.

실험에서 사용된 다중관측열의 확률 값을 구하는데 사용한 퍼지 법칙^[10]은 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 거리값이 작은 순으로 표시했을 때 몇 번째인지를 나타내고, S는 한 열에서 선택될 다중 심볼의 수를 의미하며, w_s 는 전체 확률을 1로 했을 때 관측열의 s 번째인 심볼이 가질 확률값이다.

2.3 퍼지 개념에 의한 HMM의 원리

HMM은 이중의 확률 처리로써 하나는 현재의 천이가 이루어질 천이 확률이고, 또 하나는 천이가 이루어졌을 때 유한개의 관측 대상으로부터 각 출력 심볼이 출현될 조건부 확률을 규정하는 출력 확률 밀도 함수인데, 본 연구에서는 각 출력 심볼에 퍼지 개념을 도입하여 가중치를 줌으로서 인식률의 향상은 물론 학습 데이터의 수를 줄일 뿐 아니라 또한 전 음성 부분을 동적으로 몇 개의 구간으로 나누어 줌으로써 인식 시간 및 인식률을 증진시키는 방법에 대해서 제안한다.

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

- 상태수 : N
- 전체 심볼수 : M
- 열의 관측될 심볼수 : S
- 상태집합 $Q = \{q_1, q_2, \dots, q_N\}$
- 심볼집합 : $V = \{v_1, v_2, \dots, v_M\}$
- 관측열의 길이 : $t = 1, 2, \dots, T$

t 번째 관측 심볼열이 상태 q_i 에 있고, t+1 번째 관측 심볼열이 상태 q_j 를 선택할 확률을 A라 하면, A는 다음과 같다.

$$A = \{a_{ij}\}, \quad a_{ij} = pr(q_j \text{ at } t+1 | q_i \text{ at } t), \quad (1 \leq i, j \leq N) \quad (2)$$

t 번째 관측 심볼열이 q_j 상태에서 다중 심볼집합 $\{v_{k_1}, v_{k_2}, \dots, v_{k_s}\}$ 을 가지고 그 때의 각각의 심볼이 가질 가중치 집합을 $\{w_1, w_2, \dots, w_s, \dots, w_S\}$ 이라 하면,

$$\sum_{s=1}^S w_s = 1 \quad (3)$$

이다. 따라서, t 번째 관측 심볼열이 q_j 상태에서 다중 심볼 $v_k = \{v_{k_s}\}$ 를 선택할 확률을 B라 하면 B는 다음과 같다.

$$B = \{b_j(k)\}$$

$$b_j(k) = \sum_{s=1}^S w_s b_{j_s}(k_s) \quad (4)$$

$$= pr(v_k \text{ at } t | q_j \text{ at } t)$$

$$= pr(\{v_{k_s}\} \text{ at } t | q_j \text{ at } t)$$

(1 ≤ j ≤ N), (1 ≤ k ≤ M), (1 ≤ s ≤ S)

초기 상태에서 상태 q_i 를 선택할 확률은 다음과 같다.

$$\pi = \{\pi_i\}, \quad \pi_i = pr\{q_i \text{ at } t=1\} \quad (5)$$

관측열은 다음과 같다.

$$O = O_1, O_2, \dots, O_t, \dots, O_T \quad (6)$$

관측 심볼의 구성은 다음과 같다.

$$O_t = \{O_{t_1}, O_{t_2}, \dots, O_{t_s}\} \quad (7)$$

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

2.3.1 모델 학습 알고리즘

모델 학습에 앞서 생각 할 문제는 모델 파라메타 $\lambda = (A, B, w, \pi)$ 가 주어졌을 때 관측열 $O = O_1, O_2, \dots, O_t, \dots, O_T$ 의 계산 문제이다. 가장 단순하게 계산하는 방법으로는 가능한 모든 상태열에 대하여 A, B 행렬을 이용하여 확률을 계산하는 것인데, 능률적인 계산 방법으로 L. E. Baum에 의해 제안된 전향-후향 알고리즘⁶⁾은 다음과 같다.

먼저 전향 변수 $\alpha_i(i)$ 은

$$\alpha_i(i) = pr(O_1, \dots, O_t, i_t = q_i | \lambda) \quad (8)$$

이고, ($O_t = \{O_{t_1}, O_{t_2}, \dots, O_{t_s}\}$)로 정의 한다. 이는 주어진 모델 λ 에 대해서 시간 t까지의 다중 관측열이

O_1, O_2, \dots, O_t 로 표현되고, $O_t = \{O_{t_1}, O_{t_2}, \dots, O_{t_s}\}$ 의 다중 심볼로 구성되는데, 이때 상태가 q_i 의 확률을 가지며 다음과 같은 절차에 의해서 구할 수 있다.

단계 1. 초기화는 다음과 같다.

$$\alpha_1(i) = \pi_i b_i(O_1) = \sum_{s=1}^S \pi_i w_s b_{i_s}(O_{1_s}), \quad 1 \leq i \leq N \quad (9)$$

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

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

단계 3. 다음과 같은 결과가 나온다.

$$p(O|\lambda) = \sum_{i=0}^N \alpha_T(i) \quad (11)$$

같은 방법으로 후향 변수를 살펴 보면.

$$\beta_t(i) = pr(O_{t+1}, O_{t+2}, \dots, O_{T-1}, O_T | i_t = q_i, \lambda) \quad (12)$$

이고, $(O_t = \{O_{t_1}, O_{t_2}, \dots, O_{t_s}\})$ 로 정의할 경우 다음 절차에 의해 구할 수 있다.

단계 1. 초기화는 다음과 같다.

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

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

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

이다. 일반적인 HMM에서 초기 파라메타들로 부터 $pr(O|\lambda)$ 를 최대로 하는 $\lambda = (A, B, w, \pi)$ 를 재추정하는 것으로 Baum-Welch 재추정 알고리즘^[6]을 이용하는데, Baum-Welch의 재추정 알고리즘도 전향-후향 알고리즘과 마찬가지로 일반적인 알고리즘에서 수정하여 사용할 수

있다. 주어진 관측열과 모델 λ 에 대해서 시간 t 에서의 상태가 q_i 이고 시간 $t+1$ 에서의 상태가 q_j 일 확률 $\xi_t(i, j)$ 를 다음과 같이 정의한다.

$$\begin{aligned} \xi_t(i, j) &= pr(i_t = q_i, i_{t+1} = q_j | O, \lambda) \\ &= \frac{\sum_{s=1}^S \alpha_t(i) a_{ij} w_s b_{j_s}(O_{(t+1)_s}) \beta_{t+1}(j)}{pr(O|\lambda)} \end{aligned} \quad (15)$$

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

$$\gamma_t(i) = pr(i_t = q_i | O, \lambda) \quad (16)$$

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

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

$$\begin{aligned} \hat{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^T \gamma_t(i)} \\ &= \frac{\sum_{t=1}^{T-1} \sum_{s=1}^S \alpha_t(i) a_{ij} w_s b_{j_s}(O_{(t+1)_s}) \beta_{t+1}(j)}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)} \end{aligned} \quad (18)$$

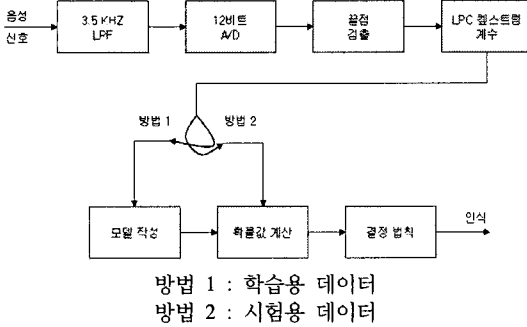
$$\hat{b}_j(k) = \frac{\sum_{t=1, O_t = v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} = \frac{\sum_{t=1, O_t = v_k}^T \alpha_t(j) \beta_t(j)}{\sum_{t=1}^T \alpha_t(j) \beta_t(j)} \quad (19)$$

2.3.2 인식 알고리즘

인식에 사용되는 알고리즘으로는 전향 알고리즘과 후향 알고리즘 외에 Viterbi 알고리즘^[6]이 있다. 그러나, 본 연구에서는 논문 [7]에 실험 결과에 의해 인식 알고리즘으로 전향 알고리즘을 사용하였다. 전향 알고리즘에 대한 수식은 식(9), (10) 및 (11)로 앞에서 표시하였다.

3. 인식 실험 결과

본 연구에서 실험에 사용된 음성 데이터 베이스는 기차역명 201개의 단어(중복된 역명 제외)를 남성화자 4명에 의해서 발음하게 하였다. 따라서, 총 인식 대상어수는 201개이다. 먼저 단독어 실험을 위해 4명 중 2명의 화자가 기차역명을 각 15회 발음한 6,030개 단어로 모델을 학습하였으며 인식은 나머지 2명의 화자가 2회씩 발음한 것을 화자 독립의 인식 실험에 사용하였다.



[그림 2] 인식 시스템 구성도

3.1 인식 실험

제안된 DMS 모델과 퍼지 개념을 이용한 HMM 방법 및 다른 여러 방법들의 인식 실험 결과를 표 1에 나타내었다.

여기서, 제안된 DMS 모델과 퍼지 개념을 이용한 HMM에 의한 인식 실험은 구간을 2, 3 및 4까지 나누어서 실험하였다. 실험 결과는 표 2와 같으며, 표 1에는 인식이 가장 좋은 것을 나타내었다.

[표 1] 단독어 인식 (단위 : %)

노선 \ 화자	화자 1	화자 2	전 체
제안된방법	93.53	94.03	93.78
DMS 모델을 이용한 HMM	90.54	92.53	91.54
MSVQ를 이용한 HMM	86.07	89.05	87.56
일반적인 HMM	79.85	83.58	81.72
DMS 모델을 이용한 DMSVQ	82.09	84.08	83.08
MSVQ	74.88	77.36	76.12
DP	77.36	80.60	78.98

[표 2] 구간수에 의한 단독어 인식 (단위 : %)

구간 \ 화자	화자 1	화자 2	전 체
2구간으로 나눈 경우	93.03	94.03	93.53
3구간으로 나눈 경우	93.53	94.03	93.78
4구간으로 나눈 경우	90.55	92.04	91.29

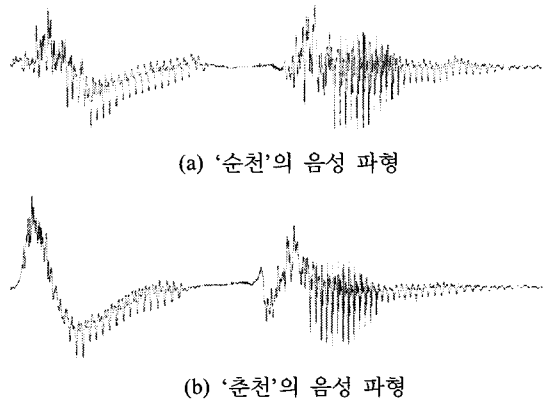
3.2 고찰

지금까지의 실험 결과에 의하면, 표 1에서는 DP 인식 방법보다 MSVQ 인식 방법이 인식 시간은 훨씬 단축되거나 인식률이 낮았다. 그러나, DMS 모델을 이용한 DMSVQ를 이용한 인식 방법에서는 인식 시간은 MSVQ나 마찬가지로인데 인식률이 약 7% 정도 높아짐을 알 수 있으며, DMS 모델을 이용한 DMSVQ 인식 방법은 일반적인 HMM 인식 방법보다도 인식률이 높음을 알 수 있다.

또한, HMM 인식 방법에서도 일반적인 HMM 보다 MSVQ를 이용한 HMM이 약 6% 정도 인식률이 높고, 마찬가지로 DMS 모델을 이용한 HMM 인식 방법은 약 10% 정도 인식률이 높아짐을 알 수 있다. 더구나, 본 연구에서 제안하는 DMS 모델과 퍼지 개념을 이용한 방법에서는 약 93.78%로 인식률이 상승함을 알 수 있다.

표 2에서는 본 연구에서 제안한 방법에 있어서 구간수를 2에서부터 4까지 실험한 것으로 구간수를 3으로 했을 때 가장 인식률이 높을 것을 알 수 있다.

여기서 오인식이 되었던 ‘순천’과 ‘춘천’의 음성 파형을 [그림 3]에 나타내었다.



[그림 3] 음성 파형

본 연구에서 제안한 방법은 동적으로 나누어준 구간마

다 각각 퍼지 개념을 이용한 다중 관측열을 사용함으로써 일반적인 HMM 인식 방법보다 약 11.5%의 인식률 증가를 보임을 알 수 있었다.

4. 결론

본 연구에서는 DMS 모델과 퍼지 개념을 이용한 HMM 모델을 제안하고, 이 모델을 이용하여 화자 독립의 단독어 인식 실험을 수행하였고 일반적으로 많이 사용되어 오고 있는 여러 방법들과 같은 조건하에서 실행하여 비교하였다.

제안된 인식 실험 방법은 각 구간을 퍼지 개념에 의한 다중 관측열을 이용함으로써 인식률이 다른 방법뿐만 아니라 일반적인 HMM 보다도 약 12% 정도 높은 약 93.78%의 인식률을 나타냄을 알 수 있었다.

오인식된 단어들을 살펴보면, 비슷한 단어들 간의 애매한 발음(예를 들면, '대천'을 '대전'에 가깝게, '순천'을 '춘천'에 가깝게 등)과 잡음이 심한 경우로 D/A에 의해 들어 보아도 구별이 가지 않는 경우가 많았다.

따라서, 화자가 정확하게 발음해야 함은 물론 잡음 속에서도 시작점과 끝점을 잘 잡아주며 시스템에 잘 적응하고 보다 더 나은 인식률을 얻을 수 있는 인식 시스템의 개발이 필요하고, 한국어의 음운학적인, 그리고 음향학적인 특성에 맞는 알고리즘의 개발이 필요하다.

참고 문헌

[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, Vol. 1, pp. 4-29, Apr. 1984

[3] D. K. Burton, J. E. Shore and J. T. Buck, "Isolated-Word Speech Recognition using Multisection Vector Quantization Codebooks", IEEE Trans. of Acoustics, Speech, Signal Processing, Vol. ASSP-33, No. 4, Aug. 1985.

[4] Tae Ock Ann and Sun hyub Kim, "An automatic Speech Recognition of Computer Using Time Sequential Vector Quantization", The Institute of Electronics Engineers of Korea, Vol. 27, No. 7, July. 1990.

[5] Tae Ock Ann and Young Kyu Byun, "A Study on Speech Recognition using DMS Model", The Acoustical Society of Korea, Vol. 13, No. 2E, pp. 41-50, Dec. 1994.

[6] L. R. Rabiner and B. H. Juang, "An Intorduction to Hidden Markov Models", IEEE ASSP Magazine, JAN. 1986.

[7] T. O. Ann, Y. G. Byun and S. H. Kim, "Korean Speech Recognition using DHMM", The Acoustical Society of Korea, Vol. 10. No. 1, pp. 52-61, Feb. 1991.

[8] 안태옥, 변용규, 김순협, "MSVQ를 이용한 HMM에 의한 단독어 인식", 대한전자 공학회, 제 27권 제 9 호, pp. 158-165, Sep. 1990.

[9] 안태옥, "Speech Recognition using MSHMM based on Fuzzy Concept", 한국 음향학회지, 제16권 2E호, pp. 55-61, Sep. 1997.

안 태 옥(Tae-Ock Ann)

[정회원]



- 1992년 2월 : 광운대학교 대학원 컴퓨터공학과 (공학박사)
- 현재 : 호원대학교 컴퓨터·게임학부 교수

<관심분야>

신호처리, 패턴인식, 음성인식 등