

FSVQ, 퍼지 개념 및 이중 스펙트럼 특징을 이용한 HMM에 기초를 둔 음성 인식 (HMM-based Speech Recognition using FSVQ, Fuzzy Concept and Doubly Spectral Feature)

정의봉(Eui-Bung Jung)¹⁾

요 약

본 논문은 화자 독립의 단독어 인식에 관한 연구로써, FSVQ(first section vector quantization), 퍼지 이론 및 이중 스펙트럼 특징을 이용한 HMM(hidden Markov model) 모델을 제안한다.

제안된 연구 방법에서, 이중 특징 파라메타로써 LPC 켈스트럼과 LPC 켈스트럼의 회귀 계수를 사용한다. 학습 데이터는 몇 개의 구간으로 나누어지며, 첫 번째 구간의 코드북(codebook)을 만든 후, 첫번째 구간의 코드북으로부터, 퍼지 개념을 도입하여 확률값이 큰 순서에 의해 다중 관측열을 구한다. 그 다음, 첫 번째 구간의 관측열을 학습시키고, 같은 방법으로 확률값을 얻은 단어가 인식되어진다.

제안된 방법에 의한 인식 실험을 수행하는 것 이외에도 비교를 위하여 다른 방법의 인식 실험을 같은 조건하에서 같은 데이터로 수행하였다.

실험 결과, 본 연구에서 제안한 방법이 다른 방법들보다 인식률이 우수함을 입증하였다.

ABSTRACT

In this paper, we propose a HMM model using FSVQ(First Section VQ), fuzzy theory and doubly spectral feature, as study on the isolated word recognition system of speaker-independent.

In the proposed paper, LPC cepstrum coefficients and regression coefficients of LPC cepstrum as doubly spectral feature be used. And, training data are divided several section and first section is generated codebook of VQ, and then is obtained multi-observation sequences by order of large propabilistic values based on fuzzy rule from the codebook of the first section. Thereafter, this observation sequences of first section is trained and is recognized a word to be obtained highest probaility by same concept.

Besides the speech recognition experiments of proposed method, we experiment the other methods under the equivalent environment of data and conditions.

In the whole experiment, it is proved that the proposed method is superior to the others in recognition rate.

Keywords : FSVQ, Fuzzy rule, Doubly spectral feature

논문접수 : 2004. 4. 10.

심사완료 : 2004. 4. 16.

1) 정회원 : 호원대학교 디지털영상정보학부

본 연구는 호원대학교 교내학술연구조성비에 의해 연구 되었습

I. 서론

인간은 기계를 만들고 편의를 위해 여러 가지 기능을 추가하여왔다. 이에 따라 작동법을 모르는 사람에 있어서는 기계를 작동하는데 더 큰 어려움이 뒤따르게 되었는데 음성을 이용하면 이런 어려움을 극복할 수 있다. 이와 같이 음성을 정보 교환의 수단으로 사용하려면 기계가 음성을 인식하는 것이 선행되어야 하는데, 음성 인식이란 인간의 음성을 기계가 알아듣게 하는 것이다. 이에 반해, 음성 합성이란 기계가 사람처럼 말을 할 수 있게 하는 것을 의미한다.

본 논문은 중규모 어휘의 화자 독립의 단독어 인식에 관한 연구를 행하는 것이다. 그런데, 음성 인식을 하는데 있어서 중규모 이하의 단어에 있어서는 음소나 음절 단위와 같이 써브워드(subword) 단위로 음성 인식을 행하는 것보다 단어 단위로 음성 인식을 행하는 것이 더 실용적이라 할 수 있다. 이와 같이 단어 단위의 음성 인식을 하는데 있어서, 기존의 DP에 의한 인식^[1]은 인식 시간이 너무 길고 기억 용량도 크며, VQ^[2,3]나 MSVQ^[4,5]는 인식 시간 및 기억 용량은 적으나, 인식률이 떨어진다.

따라서, 다른 방법보다 인식률 및 계산 시간 등이 우수한 모델로 HMM 모델을 선정하여, 앞으로 다가올 인간과 기계와의 통신 시대에 대비하여 기계가 우리 인간의 자연어를 자연스럽게 인식할 수 있도록 하기 위한 일환으로 단독어 인식 시스템이 개발되고 있다.

그러나, 일반적인 HMM^[6-9]의 경우도 모델을 작성하는데 많은 화자의 많은 학습용 데이터가 필요하며, 모델 학습시에 참여하지 않은 화자의 음성은 인식률이 현저하게 떨어질 수 있는데 본 연구에서는 이중 특징 벡터^[8,9]로써 정적 스펙트럼 특징으로 LPC 켈스트럼 계수를 사용하고 동적 스펙트럼 특징으로는 LPC 켈스트럼의 회귀 계수를 사용하였으며, 관측열(observation sequence)을 구할 시에 퍼지 개념을 도입^[11]한 다중 관측열(multi-observation

sequence)를 이용하여 HMM 학습에 이용함으로써 인식률을 향상시키고, 또한 시간 단축과 첫 번째 구간의 시간 정보의 중요성을 이용하는 FSVQ 코드북을 작성^[12]하여, 인식시에 첫 번째 구간만 확률값을 비교하여 인식 대상 단어의 후보 수를 줄이고 인식률도 향상시키는 HMM 모델을 제안한다.

II. 제안된 시스템

일반적인 HMM에 의한 단독어 인식에 있어서는 VQ 코드북을 작성한 후 이 코드북에 근거하여 관측열을 구하는 관계로 시간이 많이 걸리며, 시간의 변화에 대한 정보를 포함하지 못하여 인식률이 낮다. 이에 MSVQ 코드북을 이용한 HMM에 의한 단독어 인식에 관한 연구가 나와서 인식률을 높이고 인식 시간도 단축시켰다.

본 연구에서는 음성의 앞부분에 높은 음성 정보가 있다는 관점에서의 FSVQ 개념에 정적인 특징 벡터만 이용하여 HMM 모델링을 행하던 것보다 동적인 특징 벡터까지를 이용한 HMM을 이용한 인식 방법이 인식률이 높다는 개념을 도입하였다. 또한, 일반적인 이산 HMM의 경우는 각 열마다 하나의 심볼만 관측되는 것으로 생각하였는데 본 연구에서는 각 열마다 몇 개의 다중 심볼이 적당한 가중치를 가지고 한 열을 구성할 수 있다는 퍼지를 이용한 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 몇 개의 관측될 심볼에 속하면 확률값을 가질 수 있으므로 인식시에 어느 정도의 확률값을 가져 인식될 가능성을 높여준다.

따라서, 본 연구에서는 첫 번째 구간에 높은 음성 정보가 있다는 개념의 FSVQ 코드북, 퍼지 개념의 다중 관측열 및 이중 특징 벡터

에 의한 HMM 음성 인식 시스템을 제안한다.

1. 기본적인 이론

FSVQ 코드북, 퍼지의 다중 관측열 및 이중특징 벡터에 의한 HMM을 이용한 음성 인식을 하는데 있어서 음성 신호를 관측열로 변환하는 데는 일반적인 HMM과 같이 먼저 음성의 전처리 과정, 특징 추출, 그리고 데이터 압축 등을 거쳐서 이루어진다. 이런 분석이 끝난 후에 본 연구에서는 구간을 나누어 첫 번째 구간의 이중 특징 벡터에 의한 코드북을 작성하고 퍼지 개념에 의한 다중관측열을 이용하여 HMM 모델 및 인식 알고리즘을 세운다.

1) 2중 스펙트럼 특징에 의한 HMM^[8,9]의 개념 도입

2중 특징 벡터에 의한 HMM이란 정적 스펙트럼 특징 파라미터와 동적 스펙트럼 특징 파라미터를 함께 모델링한 것을 말한다. 논문 [8]에 의하면 정적 특징과 동적 특징 사이의 관계가 정적 특징과 정적 특징 및 동적 특징과 동적 특징 사이의 상관관계 보다도 매우 적다는 사실을 이용하여 정적 스펙트럼과 동적 스펙트럼 특징을 바탕으로 2중 특징 벡터의 HMM을 계산의 증가 없이 모델링 할 수 있다.

2) 퍼지 개념^[11]을 이용한 다중 관측열 나열

어떤 발음을 했을 때 발음할 때마다 또는 화자에 따라 특성이 달라지지만 그 발음이 가지는 특성이 많이 벗어나지는 않을 거라는데 유의하여 퍼지 개념을 이용하였다.

즉, 각 프레임의 벡터와 VQ 코드북의 코드워드 중 거리값이 가장 짧은 것을 심볼로 선택하는 것이 일반적인 HMM에서 관측열을 구하는 방법이었다. 그런데, 일반적인 HMM에서 관측열을 구하는 것과는 달리 퍼지 개념을 이용하여 각 열마다 몇개의 유사한 특징을 가지는 심볼을 선택하는 방법을 제안한다.

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 인식시에도 심볼로 선택될 확률을 크게 가질 수 있다는 개념에서 비롯되었다.

실험에서 사용된 다중 관측열을 구하는데 사용한 퍼지 법칙은 다음과 같다.

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

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

3) FSVQ 이론^[12]

Burton^[4]의 MSVQ 코드북을 이용한 음성 인식에 따르면 MSVQ 코드북 작성시에는 시간 정보를 이용하는 관계로 단일 VQ를 이용한 음성 인식 방법보다 인식을 뿐만 아니라 인식 시간도 단축됨을 알 수 있다. 따라서, 본 연구에서는 HMM 모델에도 시간 감축 및 시간 정보를 이용하기 위해 MSVQ에 의한 코드북 작성^[10]에서와 마찬가지로 다중 구간 중 첫 번째 구간의 코드북을 작성하여 구간 수를 줄이면서 HMM 모델에 의한 인식 실험^[12]을 수행한다.

본 연구에서 제안한 FSVQ, 퍼지 개념 및 이중 특징 벡터에 의한 HMM 인식 실험은 이중 특징 벡터를 파라미터로 사용하고, 전체 단어에 대한 첫 번째 구간의 코드북은 그 단어를 몇 개의 구간으로 나누고 첫 번째 구간의 집단화 알고리즘을 이용하여 작성하였으며, 모두 128개의 코드워드로 이루어졌다. 따라서, 2구간으로 이루어진 FSVQ 코드북의 경우는 첫 번째 구간의 코드북을 만들어 인식 실험을

행하여 대상 어휘를 1/20로 줄여준 후, 단어 전체를 한 구간으로 해서 코드북을 만들어 줄여준 대상 어휘를 중심으로 최종적인 인식 실험을 행한다. 또한, 4 구간으로 이루어진 FSVQ 코드북의 경우는 첫 번째 구간의 코드북을 만들어 인식 실험을 행하여 대상 어휘를 1/10로 줄여준 후, 그 다음에는 2 구간으로 나누어준 것과 같은 방법으로 인식 실험을 행하였고, 8 구간으로 이루어진 FSVQ 코드북의 경우는 첫 번째 구간의 코드북을 만들어 인식 실험을 행하여 대상 어휘를 1/5로 줄여준 후, 그 다음에는 4 구간으로 나누어준 것과 같은 방법으로 인식 실험을 행한다.

2. 제안된 방법에 의한 음성 인식

HMM은 이중의 확률 처리로써 하나는 현재의 천이가 이루어질 천이 확률이고, 또 하나는 천이가 이루어졌을 때 유한개의 관측 대상으로부터 각 출력 심벌이 출현될 조건부 확률을 규정하는 출력 확률 밀도 함수인데, 본 연구에서는 정적 특징 벡터와 동적 특징 벡터가 서로 상관 관계가 없다는 개념에서 이중 특징 벡터를 사용하고 각 출력 심벌에 퍼지 개념을 도입하여 가중치를 줌으로써 인식률의 향상은 물론 학습 데이터의 수를 줄일 뿐 아니라 또한 전 음성 부분을 몇 개의 구간으로 나누어 첫 번째 구간에만 시간 정보를 주어 인식률을 높이는 한편 인식 시간을 줄이며, 더 나아가 구간 수를 줄일 때마다 인식 대상 후보의 수를 줄여줌으로써 인식 시간 및 인식률을 증진시키는 방법에 대해서 제안한다.

1) 이중 특징 벡터에 의한 HMM의 개념 도입

이중 특징 벡터에 의한 HMM이란 정적 스펙트럼의 특징 파라메타와 동적 스펙트럼의 특징 파라메타를 함께 모델링한 것을 말한다.

본 연구에서는 일반적인 HMM에서와 마찬가지로 이중 특징 벡터인 정적 특징 벡터와

동적 특징 벡터가 서로 상관 관계가 없다는 가정하에서 알고리즘을 적용한다. 즉, 정적 특징에 의한 코드워드 (O_{I_t})와 동적 특징에 의한 코드워드 (O_{D_t})가 서로 상관관계가 없다고 했으므로 식 (2)을 다시 쓰면,

$$b_s(O_t) = \Pr(O_t | S) = b_s^I(O_{I_t}) b_s^D(O_{D_t}) \quad (2)$$

로 된다. 여기서,

$b_s^I(O_{I_t})$ 는 상태 S에서 정적 벡터 (O_{I_t})가 나올 확률이며,

$b_s^D(O_{D_t})$ 는 상태 S에서 동적 벡터 (O_{D_t})가 나올 확률이다.

퍼지 개념을 이중 특징 벡터에 사용시에는 각 벡터마다 다중 관측열을 얻을 때 각각의 가중치 w 가 구해진다. 이 때 구해지는 가중치를 정적 특징의 경우는 $w_s^I(k_s^I)$ 라 하고, 동적 특징의 경우는 $w_s^D(k_s^D)$ 라 한다. 본 연구에서는 가중치들

$w_s(k_s) = w_s^I(k_s^I) * w_s^D(k_s^D)$ 로 적용하였는데 $w_s^I(k_s^I)$ 와 $w_s^D(k_s^D)$ 는 사실상 가중치가 같으므로 따로 생각할 여지가 없으며 또한 s번째의 가중치는 어느 경우이나 똑같으므로 가중치는 w_s 로만 표시하겠다.

2) 제안된 방법을 이용한 HMM의 원리

본 연구에서 제안하고 있는 이중 특징 벡터와 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 : $t = 1, 2, \dots, T$

t 번째 관측 심볼열이 상태 q_i 에 있고,
 $t+1$ 번째 관측 심볼열이 상태 q_j 를 선택할 확률

$$A = \{a_{ij}\}, a_{ij} = pr(q_j \text{ at } t+1 | q_i \text{ at } t), (1 \leq i, j \leq N) \quad (3)$$

t 번째 관측 심볼열이 q_j 상태에서 다중 심볼 집합 $\{v_{k1}, v_{k2}, \dots, v_{ks}, \dots, v_{kS}\}$ 을 가지고 그 때의 각각의 심볼이 가질 가중치 집합을 $\{w_1, w_2, \dots, w_s, \dots, w_S\}$ 이라 하면

$$\sum_{s=1}^S w_s = 1 \quad (4)$$

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

$$\begin{aligned} B &= \{b_j(k)\}, \\ b_j(k) &= \sum_{s=1}^S w_s b_{js}(k_s) \quad (5) \\ &= pr(v_k \text{ at } t | q_j \text{ at } t) \\ &= pr(\{v_{ks}\} \text{ at } t | q_j \text{ at } t) \end{aligned}$$

$(1 \leq j \leq N), (1 \leq k \leq M), (1 \leq s \leq S)$

초기 상태에서 상태 q_i 를 선택할 확률

$$\pi = \{\pi_i\}, \quad \pi_i = pr\{q_i \text{ at } t=1\} \quad (6)$$

관측열

$$O = O_1, O_2, \dots, O_t, \dots, O_T \quad (7)$$

관측 심볼의 구성

$$O_t = \{O_{t1}, O_{t2}, \dots, O_{tS}\} \quad (8)$$

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.

(1) 모델 학습 알고리즘

모델 학습에 앞서 생각 할 문제는 모델 파라메타 $\lambda = (A, B, w, \pi)$ 가 주어졌을 때 관측열 $O = O_1, O_2, \dots, O_T$ 의 계산 문제이다. 가장 단순하게 계산하는 방법으로는 가능한 모든 상태열에 대하여 A, B 행렬을 이용하여 확률을 계산하는 것인데, 능률적인 계산방법으로 L. E. Baum에 의해 제안된 전향-후향 알고리즘^[6]은 다음과 같다.

먼저 전향 변수 $\alpha_t(i)$ 은

$$\alpha_t(i) = pr(O_1, O_2, \dots, O_t, i_t = q_i | \lambda) \quad (9)$$

$$(O_t = \{O_{t1}, O_{t2}, \dots, O_{tS}\})$$

로 정의 한다.

이는 주어진 모델 λ 에 대해서 시간 t 까지의 다중 관측열이 O_1, O_2, \dots, O_t 로

표현되고, 각각의 O_t 는

$\{O_{1s}, O_{2s}, \dots, O_{ts}\}$ 의 다중 심볼로 구성되는데, 이때 상태가 q_i 의 확률을 가지며 다음과 같은 절차에 의해서 구할 수 있다.

단계 1. 초기화

$$\alpha_1(i) = \pi_i b_i(O_{1s}) = \sum_{j=1}^S \pi_j w_s b'_{js}(O_{1s}) b^D_{js}(O_{D1s}), 1 \leq i \leq N \quad (10)$$

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

$$\alpha_{t+1}(j) = \sum_{i=1}^N \left[\sum_{k=1}^N \alpha_t(i) a_{ik} \right] w_s b'_{js}(O_{t(t+1)s}) b^D_{js}(O_{D(t+1)s}) \quad (11)$$

단계 3. 그러면

$$p(O | \lambda) = \sum_{i=0}^N \alpha_{T(i)} \quad (12)$$

같은 방법으로 후향 변수를 살펴 보면

$$\beta_t(i) = pr(O_{t+1}, O_{t+2}, \dots, O_{T-1}, O_T | i_t = q_i, \lambda) \quad (13)$$

$(O_t = \{O_{1s}, O_{2s}, \dots, O_{ts}\})$

로 정의할 경우 다음 절차에 의해 구할 수 있다.

단계 1. 초기화

$$\beta_T(i) = 1, 1 \leq i \leq N \quad (14)$$

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

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (15)$$

$$= \sum_{i=0}^N \sum_{j=1}^S a_{ij} w_s b'_{js}(O_{t(t+1)s}) b^D_{js}(O_{D(t+1)s}) \beta_{t+1}(j)$$

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 이용하는 데, 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.

$$\xi_t(i, j) = pr(i_t = q_i, i_{t+1} = q_j | O, \lambda) \quad (16)$$

$$= \frac{\sum_{i=1}^S \alpha_t(i) a_{ij} w_s b'_{js}(O_{t(t+1)s}) b^D_{js}(O_{D(t+1)s}) \beta_{t+1}(j)}{pr(O | \lambda)}$$

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

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

여기서 현재 모델

$\lambda = (A, B, w, \pi)$ 이라 하고, 다시 추

정된 모델은 $\lambda = (A, B, w, \pi)$ 이라 하면 새로운 모델 파라메타는 아래의 식들과 같다.

$$\pi_i = \gamma_1(i), 1 \leq i \leq N \quad (18)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^T \gamma_t(i)} \quad (19)$$

$$\begin{aligned}
 &= \frac{\sum_{i=1}^{T-1} \sum_{j=1}^S \alpha_i(j) a_{ij} w_s b'_{js}(O_{R(t+1)s}) b^D_{js}(O_{D(t+1)s}) \beta_{t+1}(j)}{\sum_{i=1}^T \alpha_i(j) \beta_i(j)} \\
 b_j(k) &= \frac{\sum_{i=1, i \neq v_k}^T \gamma_i(j)}{\sum_{i=1}^T \gamma_i(j)} = \frac{\sum_{i=1, i \neq v_k}^T \alpha_i(j) \beta_i(j)}{\sum_{i=1}^T \alpha_i(j) \beta_i(j)} \quad (20)
 \end{aligned}$$

(2) 인식 알고리즘

인식에 사용되는 알고리즘으로는 전향 알고리즘과 후향 알고리즘 외에 Viterbi 알고리즘^[6]이 있다. 그러나, 본 연구에서는 논문 [10]에 실험 결과에 의해 전향 알고리즘이 인식률이 Viterbi 알고리즘보다 좋았으므로 인식 알고리즘으로 전향 알고리즘을 사용하였다. 따라서 전향 알고리즘을 살펴보면 다음과 같다.

단계 1. 초기화

$$\alpha_i(i) = \pi_i b_i(O_{i1}) = \sum_{s=1}^S \pi_i w_s b'_{is}(O_{Nis}) b^D_{is}(O_{Dis}), \quad 1 \leq i \leq N \quad (21)$$

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

$$\alpha_{t+1}(j) = \sum_{i=1}^S \left[\sum_{i=1}^N \alpha_i(i) a_{ij} \right] w_s b'_{js}(O_{R(t+1)s}) b^D_{js}(O_{D(t+1)s}) \quad (22)$$

단계 3. 그러면

$$p(O | \lambda) = \sum_{i=0}^N \alpha_T(i) \quad (23)$$

3) 제안된 방법에 의한 모델 작성 및 인식 시스템의 구조.

일반적인 HMM이 VQ 코드북에 의해 관측열을 구하는데 비해 FSVQ, 퍼지 개념 및 이

중 특징 벡터를 이용한 HMM은 다중 구간으로 구간을 나눈 후 첫 번째 구간의 이중 벡터에 의한 코드북을 구한 뒤, 학습용 데이터들도 같은 수의 다중 구간으로 나누어주어서 첫 번째 구간만 HMM 학습을 시키는 것이다. 여기서, 관측열은 퍼지 개념을 도입하여 다중 관측열을 이용한다. 이에 대한 순서도는 그림 1에 나타내었다. 또한 인식시에도 마찬가지로 같은 수의 다중 구간으로 나누어 준 후 첫 번째 구간에 대해서만 인식을 시킨다. 이 때, 첫 번째 구간만 확률값을 측정할 수 있는 관계로 구간수를 줄일 때마다 후보수를 줄여줌으로써 인식 시간을 줄일 수 있다. 이에 대한 순서도는 그림 2에 나타내었다.

III. 인식 실험 결과

본 연구에서 실험에 사용된 음성 데이터 베이스는 기차역명 201개의 단어(중복된 역명은 하나만 사용)를 남성화자 4명에 의해서 발음하게 하였다. 따라서, 총 인식 대상 단어수는 201개이다. 먼저 단독어 실험을 위해 4명 중 2명의 화자가 기차역명을 각각 5회 발음한 2,010개의 keyword를 가지고 학습하였으며 인식은 나머지 2명이 2회씩 발음한 것을 화자 독립 실험에 사용하였다.

3.1 인식 시스템 구성

본 연구에서 실험한 음성 인식 시스템은 그림 3과 같다.

실험에 사용된 모든 기차역명 데이터는 잡음이 섞인 환경하에서 릴 테이프에 의해 입력된 음성 신호로써, 샘플링 주파수를 8 KHz로 하였으며, 3.5 KHz의 저역 여파기를 통과한 후 12 Bit A/D 변환기를 거쳐 음성 데이터를 구하였다. 그리고, 시작점과 끝점을 검출한 후, LPC 계수보다 LPC 켈스트럼 계수를 특징 벡터로 사용하는 것이 더 좋다는 고찰^[13]에 따라

특징 파라메타로는 LPC 켈스트럼과 그의 회귀계수를 사용하였으며, 이것을 이용하여 각 단어의 모델을 구하고, 이 모델을 기초로 인식 실험을 행했다.

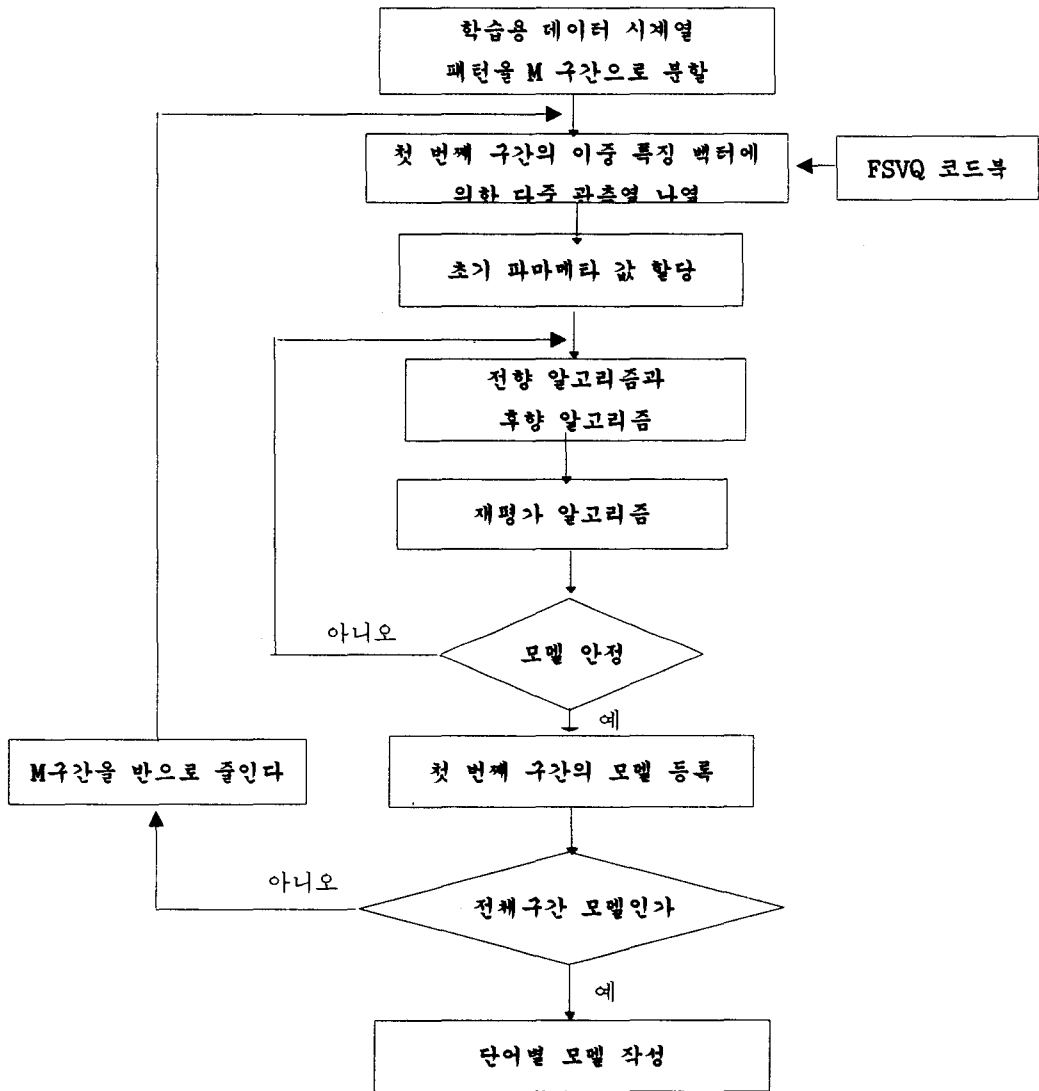


그림 1. 제안된 방법의 모델 작성
 Fig. 1. The model generation of proposed method

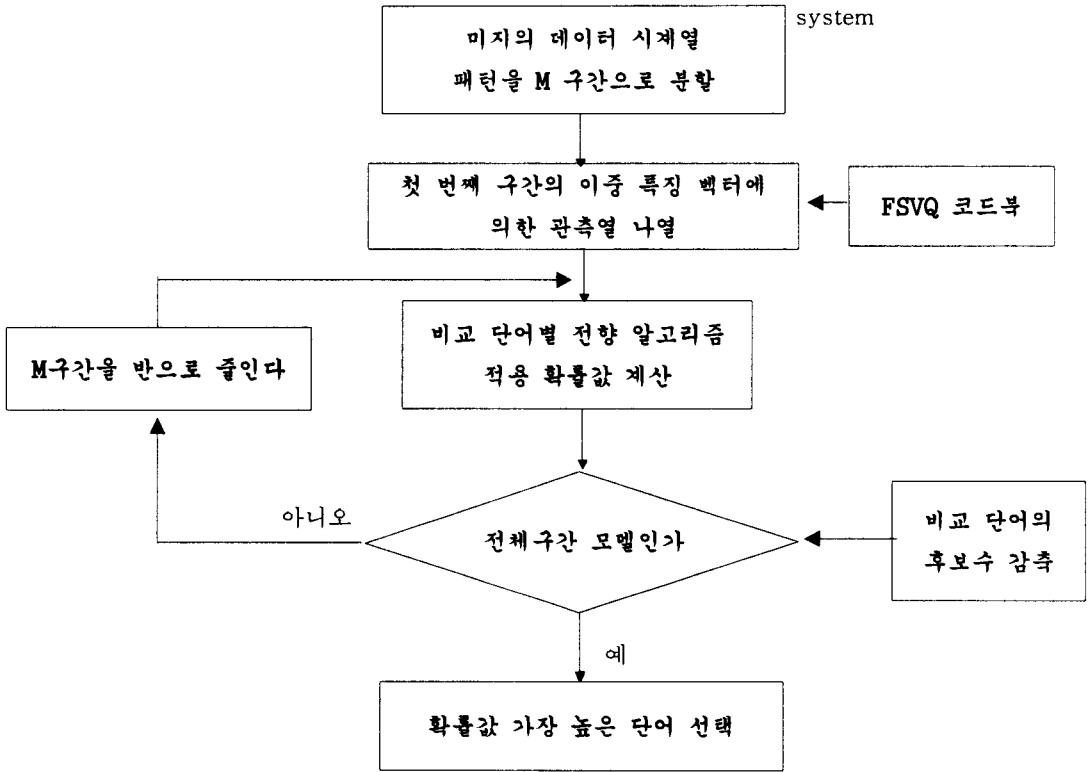


그림 2. 제안된 인식 방법

Fig. 2. The proposed recognition method.recognition system

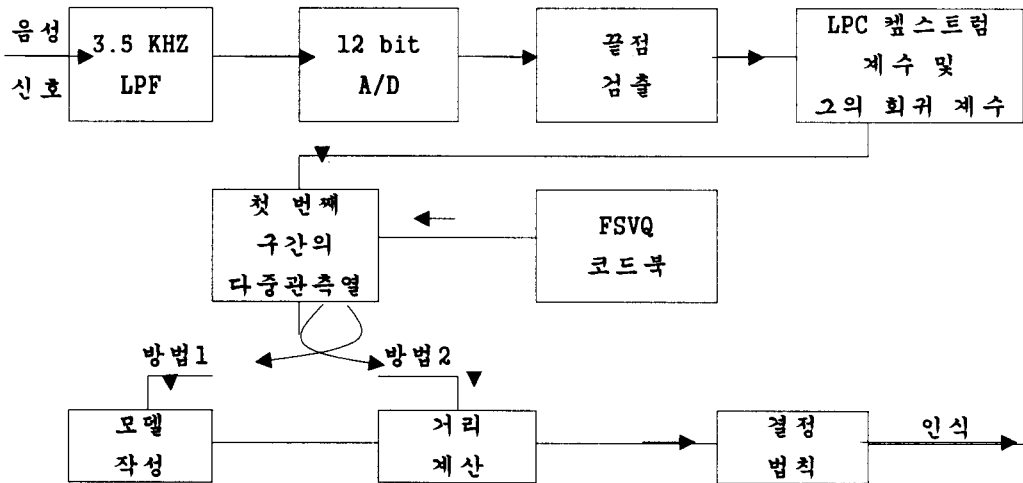


그림 3. 인식 시스템의 구성도

Fig 3. Block diagram of recognition

방법 1 : 학습용 데이터

방법 2 : 시험용 데이터

2. 인식 실험

학습데이터는 전국 기차역명 201개의 단어를 남성화자 4명 중 2명에 의해서 각각 5회씩 발음한 2,010개의 단어를 사용하였다. 단독어 인식 실험은 화자 독립으로 나머지 2명이 2번씩 발음한 것으로 실험을 수행하였다. 제안된 FSVQ, 퍼지 개념 및 이중 특징 벡터를 이용한 HMM 방법 및 다른 여러 방법들의 결과를 표 1에 나타내었다.

여기서, FSVQ에 기초를 둔 HMM의 경우는 구간을 2로 나누었을 때 인식률이 가장 좋았으므로 2 구간으로 나누어서 실험하였으며, FSVQ, 퍼지 개념 및 이중 특징 벡터를 이용한 HMM에 의한 인식 실험은 구간을 2, 4 및 8까지 나누어서 실험하였다. 실험 결과는 표 2와 같으며, 표 1에는 인식률이 가장 좋은 것을 나타내었다.

표 1. 단독어 인식

Table 1. Isolated word recognition (단위 : %)

화자 노선	화자 1	화자 2	전 체
제안된 HMM	93.53	95.02	94.28
FSVQ와 퍼지 개념을 이용한 HMM	92.04	94.03	93.03
FSVQ와 이중 스펙트럼 특징을 이용한 HMM	90.55	92.04	91.29
FSVQ를 이용한 HMM	88.06	90.05	89.30
MSVQ를 이용한 HMM	86.07	89.05	87.56
일반적인 HMM	79.85	83.58	81.72
MSVQ	74.88	77.36	76.12
DP	77.36	80.60	78.98

표 2. 구간수에 의한 단독어 인식
Table 2. Isolated word recognition by
number of sections

(단위 : %)

화자 구간	화자 1	화자 2	전 체
2구간으로 나눈 경우	93.53	95.02	94.28
4구간으로 나눈 경우	93.03	94.03	93.53
8구간으로 나눈 경우	89.30	91.29	90.30

3. 고찰

지금까지의 실험 결과에 의하면, FSVQ, 퍼지 개념 및 이중 특징 벡터를 이용한 HMM에 의한 인식 방법이 인식률이 좋은 것을 알 수 있다.

본 연구에서의 개념은 FSVQ에다 퍼지 개념과 이중 특징 벡터를 이용함으로써 HMM 음성 인식 방법을 개선시킨 방법으로 정적 특징 벡터에다 동적 특징 벡터를 포함시킨 이중 특징 벡터를 사용함으로써 음성의 많은 특성을 잡아내고 퍼지 이론의 다중 관측열을 이용하고 다중 구간 중 첫 번째 구간이 가장 많은 정보를 가지고 있다는 개념에서 연구함으로써, 표 1에서 보는 바와 같은 인식률을 얻었다.

또한, 제안된 방법에 있어서도 표 2에서와 같이 구간수에 따라 인식률이 달랐는데 구간수를 2로 한 경우가 4로 한 경우보다 약간 인식률이 높았으며, 따라서, 시간 측면에서는 구간수를 4로 한 경우가 더 처리 속도가 감소된다. 여기서, 구간수를 8로 한 경우는 우리말의 대부분 음절수가 2 내지는 3인데 너무 많이 구간수를 나눔으로써 인식률이 떨어짐을 알 수 있다.

IV. 결론

HMM 모델을 이용한 음성 인식을 하는데 있어서 모델의 파라미터를 잘 학습시키는 것은 중요한 일이다. 따라서, 본 연구에서는 FSVQ, 퍼지 개념 및 이중 스펙트럼 특징을

이용한 HMM 모델을 제안하고, 이 모델을 이용하여 화자 독립의 단어 인식 실험을 수행하였고 일반적으로 많이 사용되어 오고 있는 DP 방법과 MSVQ 방법에 의한 인식 실험 및 일반적인 HMM, MSVQ를 이용한 HMM과 FSVQ를 이용한 HMM 인식 방법과 같은 조건하에서 인식 실험을 수행하여 비교하였다.

그런데, 본 연구에서처럼 이중^중 특징 벡터를 사용하고 퍼지 개념을 이용하고 시간정보와 단어의 앞부분에 많은 음성 정보가 있다는 개념을 도입함으로써, 단일 관측열과 FSVQ 코드북을 이용하여 첫번째 구간만 관측열을 구하고 모델의 구간 수를 줄이면서 인식 후보 단어수를 줄여준 인식 실험보다 인식률을 향상시켰다.

오인식된 단어들을 살펴보면, 비슷한 단어 들간의 애매한 발음과 잡음이 심한 경우로 D/A 컨버팅 시켜 들어보아도 구별이 가지 않는 경우였다.

따라서, 화자가 정확하게 발음해야 함은 물론 잡음에 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] F. K. Soong, A. E. Rosenberg, L. R. Rabiner and B. H. Juang, "A Vector Quantization Approach to Speaker Recognition", IEEE Trans. of Acoustics, Speech Signal Processing, Vol. ASSP-33, No. 4, Oct. 1985.
- [3] D. K. Burton and J. E. Shore, "Speaker-Dependent Isolated Word Recognition using Speaker-Independent Vector Quantization Codebooks Augmented with Speaker-Specific Data", IEEE Trans. of Acoustics, Speech, and Signal Processing. ASSP-33, No. 2. pp 440-443. Apr. 1985.
- [4] 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.
- [5] 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.
- [6] L. R. Rabiner and B. H. Juang, " An Introduction to Hidden Markov Models", IEEE ASSP Magazine, JAN. 1986.

[7] 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.

[8] Furui S., "Speaker-Independent Isolated Word Recognition using Dynamic Features of Speech Spectrum" IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-34(1) pp.52-59, Feb. 1986.

[9] M. Niishmura, "HMM-based Speech Recognition using Dynamic Spectral Feature," Proc. ICASSP 89 pp 298-301, May 1989.

[10] 안태욱, 변용규, 김순협, "MSVQ를 이용한 HMM에 의한 단독어 인식", 대한전자공학회지, 제 27권, 제 9호, pp. 158-165, 1990. 9.

[11] 안태욱, 김순협, "퍼지에 기초를 둔 HMM을 이용한 음성 인식", 대한전자공학회지, 제 28권, B편 제 12호, pp. 68-74, 1991. 12.

[12] 안태욱, "FSVQ에 기초를 둔 HMM을 이용한 음성 인식", 한국컴퓨터산업교육학회, 제 2권, 제 5호, pp. 661-670, 2001. 5.

[13] 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.

정의봉



1986. 광운대학교 대학원 컴퓨터 공학과 졸업 (공학석사)

1992. 건국대학교 대학원 전자공학과 졸업 (공학박사)

관심분야: 디지털영상정보처리, 멀티미디어 CAD, 음성인식 및 합성, 인공지능