

DMS 모델과 이중 스펙트럼 특징을 이용한 HMM에 의한 음성 인식

안태옥^{1*}

HMM-based Speech Recognition using DMS Model and Double Spectral Feature

Tae-Ock Ann^{1*}

요약 본 논문은 화자 독립의 음성인식을 위한 연구로써, DMS 모델에 의한 DMSVQ(Dynamic Multi-Section Vector Quantization) 코드북과 이중 스펙트럼 특징을 이용한 HMM(Hidden Markov Model) 음성인식 방법을 제안한다. 정적 스펙트럼 특징으로서는 LPC 켈스트럼 계수를 이용하였고, 동적 스펙트럼 특징으로는 LPC 켈스트럼의 회귀계수를 사용하였다. 이들 두개의 스펙트럼 특징들을 각각 VQ 코드북으로 양자화되고, DMS 모델을 이용한 HMM은 입력으로써 정적 스펙트럼 특징과 동적 스펙트럼 특징을 받아드림으로써 모델링된다. 제안된 방법에 의한 인식 실험은 기존의 다양한 인식 방법에 의한 인식 실험들과 비교를 위해 동일한 데이터와 조건 하에서 수행하였다. 실험 결과, 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 model and double spectral feature, as a method on the speech recognition of speaker-independent.

LPC cepstrum parameter is used as a instantaneous spectral feature and LPC cepstrum's regression coefficient is used as a dynamic spectral feature. These two spectral features are quantized as each VQ codebook. HMM using DMS model is modeled by receiving instantaneous spectral feature and dynamic spectral feature by input. Other experiments to compare with the results of recognition experiments using proposed method are implemented by the various conventional recognition methods under the equivalent environment of data and conditions. Through the experiment results, it is proved that the proposed method in this paper is superior to the conventional recognition methods.

Key words : DMS, 이중 특징, 정적 스펙트럼 특징, 동적 스펙트럼 특징

1. 서론

본 논문은 중규모 어휘의 화자 독립의 단독어 인식에 관한 연구이다. 중규모 이하의 단독어에 관한 음성 인식에 있어서는 음소나 음절 단위와 같이 서브워드(subword) 단위로 음성 인식을 하는 것보다 단어 단위로 음성 인식을 행하는 것이 더 실용적이라 할 수 있다. 이와 같이 단

어 단위의 음성 인식을 하는데 있어서, 기존의 DP에 의한 인식[1]은 인식 시간이 너무 길고 기억 용량도 크며, VQ에 의한 인식 방법[2]이나 MSVQ에 의한 인식 방법은 인식 시간 및 기억 용량은 적으나, 인식률이 떨어진다. 여기서, VQ에 의한 인식 방법보다 MSVQ에 의한 인식 방법[3-4]이 시간 정보를 가지므로 인식률에서 VQ보다는 개선되기는 하나 그래도 인식률이 높지 않다. 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 이러한 VQ를 이용한 인식 방법보다 인식률이 우수한 HMM 모델을 이용한 인식 방

본 논문은 2006학년도 호원대학교 교내 학술연구 조성비에 의해 연구되었음.

¹호원대학교 컴퓨터게임학부

*교신저자 : 안태옥(ato0000@hanmail.net)

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. 또한 첫 번째 구간이 구간 중에서 가장 중요한 구간이라는 개념에서 착안한 FSVQ 코드북을 이용한 인식 방법[9-12]이 MSVQ 코드북을 이용한 인식 방법보다 인식률이 증가한다. 따라서 본 연구에서는 MSVQ 코드북이나 FSVQ 코드북을 이용하는 것보다 DMS 모델을 이용한 음성 인식 방법이 인식률이 더 높다는 점[5, 14]과 정적 스펙트럼 특징을 이용하는 단일 벡터를 사용하는 것보다 정적 스펙트럼 특징과 동적 스펙트럼 특징을 이용하는 이중 스펙트럼 특징을 사용하는 것이 인식률이 높아진다는 연구 결과[7, 10]에 착안하여 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM 음성 인식을 행하려 한다.

본 연구에서는 DMS 모델을 이용한 HMM 모델을 학습하는데 있어 이중 특징 벡터를 사용하며, 인식할 때에도 DMSVQ와 이중 스펙트럼 특징을 이용하여 구간마다 확률값을 비교하여 인식 대상 단어의 후보수를 줄이고, 계속해서 거듭되는 구간마다 후보자 수를 줄이면서 이를 반복하여 최종적으로는 단어를 인식하는 실험을 행하는 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM 음성 인식을 제안한다.

본 논문의 구성은 1장이 서론이며, 2장에서는 제안한 시스템에 대해서 설명하고, 3장에서는 실험 결과 및 고찰을 하였으며, 4장에서 결론을 내렸다.

2. DMS 모델과 이중 스펙트럼 특징을 이용한 HMM 음성인식 시스템

일반적인 HMM에 의한 단어 인식이 있어서는 VQ 코드북을 작성한 후 이 코드북에 근거하여 관측열을 구하는 관계로 시간이 많이 걸리며, 시간의 변화에 대한 정보를 포함하지 못하여 인식률이 낮다.

따라서, 정적 스펙트럼 특징으로 LPC 켈스트럼을 사용하며, 동적 스펙트럼 특징으로 LPC 켈스트럼의 회귀 계수를 사용함으로써 단시간의 시간의 변화에 따른 시간 정보를 포함하게 하고, 몇 개의 구간으로 구간을 동적으로 잡아줌으로써 같은 특징의 정보가 같은 구간에 위치하게 함으로써 인식률을 높이는 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM 인식 방법을 사용한다.

본 연구에서 제안한 방법은 음성을 몇 개의 구간으로 나누어서 첫 번째 구간부터 마지막 구간까지 인식 실험을 행한다. 인식시 인식 시간을 줄이기 위해 첫 번째 구

간의 인식이 끝나면 확률값을 비교하여 높은 순으로 인식 대상 중 몇 개의 후보를 선택한다. 이것을 반복하여 최종적으로 마지막 구간의 인식실험을 행함으로 하나의 단어만 선택하여 인식된 것으로 생각한다. 이렇게 인식을 행함으로서 인식 시간을 줄이고, 인식률도 높아진다.

2.1 DMS 모델의 개념과 작성 방법

2.1.1 DMS 모델의 개념

DMS(Dynamic Multi-section) 모델은 VQ 개념을 확장시킨 MSVQ(Multi-Section VQ) 방법이 각 구간을 등간격으로 분할하여 프레임수를 동일한 길이로 잡아주는 것과 달리 구간의 경계를 동적으로 잡아 각 구간의 프레임수를 가변 길이로 잡아 주는 모델이다.

즉, DMS 모델은 비슷한 특징을 나타내는 벡터들을 한 구간에 포함시키고 서로 비슷하지 않은 특징을 나타내는 벡터들은 다른 구간에 분할하는 것으로 음성 패턴의 연속된 시계열 상에서 서로 비슷한 특징을 가지는 벡터들을 모아 한 구간으로 만들어 주는 가변 길이로 구간을 구분하는 방법이다.

MSVQ는 구간을 일정하게 나누어줌으로 유사하지 않은 벡터들이 한 구간을 이룰 때 대표 특징 벡터가 서로 유사하지 않은 특징 벡터들 사이에서 구해지는 불합리성이 있게 되는데 이를 개선한 DMS 모델은 유사한 특징을 가지는 벡터들을 한 구간으로 만들기 위해 구간을 동적으로 분할하여 대표 특징 벡터를 구함으로써 짧게 발음되는 특성까지도 대표 특징 벡터로 선택 될 수 있고, 그 구간에 지속 시간 정보를 갖도록 만든 모델이다. 따라서 단어 패턴의 연속된 대표 특징 벡터를 사용하여 음성인식을 수행함으로써 짧은 음소의 특징 벡터도 동등한 비중을 가질 수 있도록 구간을 동적으로 나누어준다는 점에서 DMS 모델이라 한다.

2.1.2 모델 작성 방법

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 하였다.

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

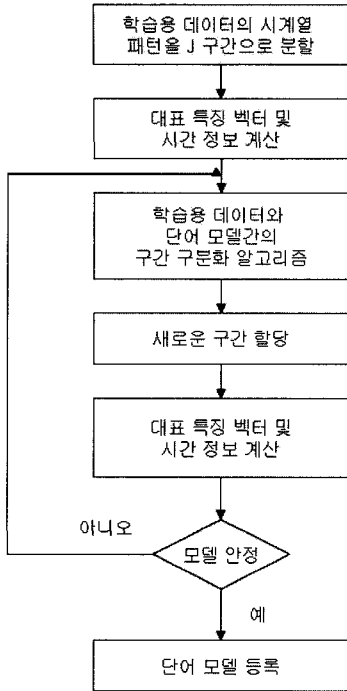


그림 1. 단어 모델 작성 방법

2.2 이중 특징 벡터에 의한 HMM의 원리

HMM은 천이들에 의해 서로 연결된 상태들의 모임으로서 각 천이에는 2종류의 확률이 관련되어 있다. 하나는 현재의 천이가 이루어질 천이 확률이고, 또 하나는 천이가 이루어졌을 때 유한개의 관측 대상으로부터 각 출력 심볼이 선택될 조건부 확률을 규정하는 출력 확률 밀도 함수이다.

일반적인 HMM에서 사용하는 기호는 다음과 같다.

상태수 : N

전체 심볼수 : M

열의 관측될 심볼수 : S

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

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

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

t 번째 관측 심볼열이 q_j 상태에서 심볼 v_k 을 선택할 확률

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

$$b_j(k) = pr(v_k | at t | q_j | at t) \quad (1 \leq j \leq N), (1 \leq k \leq M) \quad (2)$$

초기 상태에서 상태 q_i 를 선택할 확률

$$\pi = \{\pi_j\}, \quad \pi_i = pr\{q_i | at t = 1\} \quad (3)$$

관측열 $O = O_1, O_2, \dots, O_b, \dots, O_T$ (4)

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

본 연구에서는 정적 스펙트럼 특징과 동적 스펙트럼 특징의 이중 스펙트럼 특징을 이용한 HMM 인식 방법이다. 일반적인 HMM에서와 마찬가지로 2중 특징 벡터인 정적 스펙트럼 특징(LPC 켈스트럼)과 동적 스펙트럼 특징(LPC 켈스트럼의 회귀 계수)이 서로 상관관계가 없다 [7]는 가정 하에서 알고리즘을 적용한다. 또한, 정적인 의미를 가지는 것은 I(Instantaneous)라 표시하고, 동적인 의미는 D(Dynamic)로 표시하겠다. 따라서 본 연구에서는 이중 스펙트럼 특징을 모두 이용함으로

$$Pr(O_i | S) = Pr(O_{Ii}, O_{Di} | S) \quad (5)$$

으로 나타낼 수 있다.

그런데, 정적 스펙트럼 특징(O_{Ii})과 동적 스펙트럼 특징(O_{Di})은 서로 상관관계가 거의 없으므로 다음식과 같이 쓸 수 있다.

$$Pr(O_{Ii}, O_{Di} | S) \cong Pr(O_{Ii} | S) Pr(O_{Di} | S) \quad (6)$$

$$\cong Pr(O_{Ii} | M_I, M_D) Pr(O_{Di} | M_I, M_D)$$

또한, $S = (M_I, M_D)$ 라고 하면, 이것은 서로 독립이므로

$$\cong Pr(O_{Ii} | M_I) Pr(O_{Di} | M_D)$$

따라서

$$b_s(O_i) = Pr(O_i | S) = b_s^I(O_{Ii}) b_s^D(O_{Di}) \quad (7)$$

로 된다. 여기서,

$b'_s(O_{I_t})$ 는 상태 S에서 정적 스펙트럼 특징 O_{I_t} 가 나올 확률이며, $b^D_s(O_{D_t})$ 는 상태 S에서 동적 스펙트럼 특징 O_{D_t} 가 나올 확률이다.

이중 스펙트럼 특징에 의한 HMM은 정적 스펙트럼 특징과 동적 스펙트럼 특징사이의 상관관계가 없다고 가정하였으므로 파라메타의 크기에 따른 계산량의 증가 없이 모델링 될 수 있다.

전향-후향 알고리즘에서의 식을 적용시키면 다음과 같다. 먼저 전향 변수 $\alpha_i(i)$ 은

$$\alpha_i(i) = p((O_{I_1}, O_{D_1}), (O_{I_2}, O_{D_2}), \dots, (O_{I_t}, O_{D_t}), i_t = q_i | \lambda) \quad (8)$$

로 정의한다. 다음과 같은 반복에 의해서 $\Pr(O|\lambda)$ 을 구할 수 있다. 관측열은

$$\begin{aligned} O &= O_1, O_2, \dots, O_b, \dots, O_{T-1}, O_T \\ O_t &= \{O_{I_t}, O_{D_t}\} \end{aligned} \quad (9)$$

여기서, O_{I_t} 는 t 번째의 정적 스펙트럼 특징이고, O_{D_t} 는 t 번째의 동적 스펙트럼 특징이다.

단계 1. 초기화

$$\alpha_1(i) = \pi_{i_1} b'_1(O_{I_{t_1}}) b^D_1(O_{D_{t_1}}), 1 \leq i \leq N \quad (10)$$

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

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b'_j(O_{(I_{t+1})}) b^D_j(O_{(D_{t+1})}) \quad (11)$$

단계 3. 그러면

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (12)$$

같은 방법으로 후향 변수를 살펴 보면,

$$\beta_i(i) = p(O_{I_{t+1}}, O_{D_{t+1}}, \dots, O_{I_T}, O_{D_T} | i_t = q_i, \lambda) \quad (13)$$

로 정의할 경우 다음 절차에 의해 구할 수 있다.

$$\text{단계 1. 초기화 } \beta_T(i) = 1, 1 \leq i \leq N \quad (14)$$

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

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b'_j(O_{(I_{t+1})}) b^D_j(O_{(D_{t+1})}) \beta_{t+1}(j) \quad (15)$$

이다. HMM에서 초기 파라미터들로부터 $p(O|\lambda)$ 을 최대로 하는 $\lambda=(A, B, \pi)$ 을 재추정하는 것으로 Baum-Welch 재추정 알고리즘[6]을 이용한다. 여기서 현재 모델 $\lambda=(A, B, \pi)$ 이라 하고, 다시 추정된 모델은 $\hat{\lambda}=(\hat{A}, \hat{B}, \hat{\pi})$ 이라 하면 새로운 모델 파라미터는 아래의 식들과 같다.

$$\hat{\pi}_i = \gamma_1(i), 1 \leq i \leq N \quad (16)$$

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (17)$$

$$= \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b'_j(O_{(I_{t+1})}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_{t+1}(j)}$$

$$\hat{B}_j(k) = \frac{\sum_{t=1, O_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} = \frac{\sum_{t=1, O_t=v_k}^T \alpha_t(j) \beta_t(j)}{\sum_{t=1}^T \alpha_t(j) \beta_t(j)} \quad (18)$$

인식에 사용되는 알고리즘으로는 전향 알고리즘과 후향 알고리즘 외에 Viterbi 알고리즘[6]이 있다.

본 연구에서는 [논문 7]의 실험 결과에 의해 인식 알고리즘으로 전향 알고리즘을 사용하였다. 따라서 전향 알고리즘을 살펴보면 다음과 같다.

단계 1. 초기화

$$\alpha_1(i) = \pi_1 b_1(O_1) = \pi_1 b'_1(O_{I_1}) b^D_1(O_{D_1}), 1 \leq i \leq N \quad (19)$$

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

$$\alpha_{i+1}(j) = \left[\sum_{i=1}^N \alpha_i(i) a_{ij} \right] b_j^l(O_{(i+1)}) b_j^r(O_{(i+1)}) \quad (20)$$

단계 3. 그러면

$$f(O|\lambda) = \sum_{i=0}^N \alpha_i(i) \quad (21)$$

2.3 제안된 방법에 의한 모델 작성 및 인식 시스템의 구조.

일반적인 HMM이 VQ 코드북에 의해 관측열을 구하는데 비해 DMS 모델을 이용한 HMM은 이중 스펙트럼 특징을 사용함과 동시에 다중 구간으로 구간을 나누고 각 구간마다 코드북을 구한 뒤, 학습용 데이터들도 같은 수의 다중 구간으로 나누고 각 구간마다 HMM 학습을 시키는 것이다. 이에 대한 순서도는 그림 2에 나타내었다. 또한 인식시에도 마찬가지로 같은 수의 다중 구간으로 나누어 준 후 각 구간마다 인식을 시킨다. 이 때, 각 구간마다 확률값을 측정하는 관계로 어떤 구간을 인식 실험하고 다음 구간으로 인식 실험이 넘어갈 때 후보수를 줄여줌으로써 인식 시간을 줄일 수 있다. 이에 대한 순서도는 그림 3에 나타내었다.

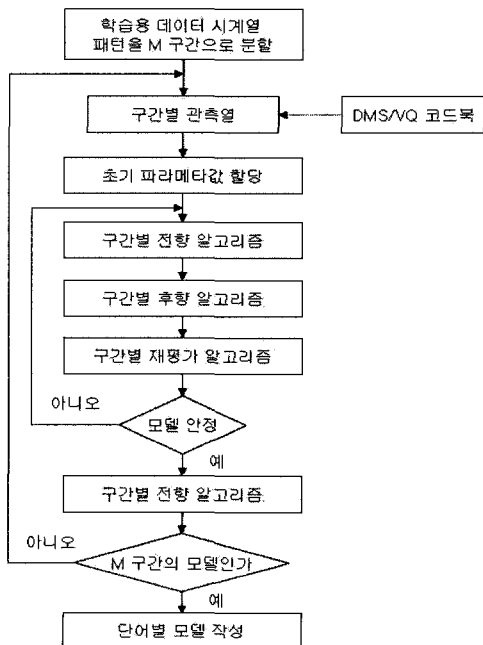


그림 2. 제안된 방법의 모델 작성

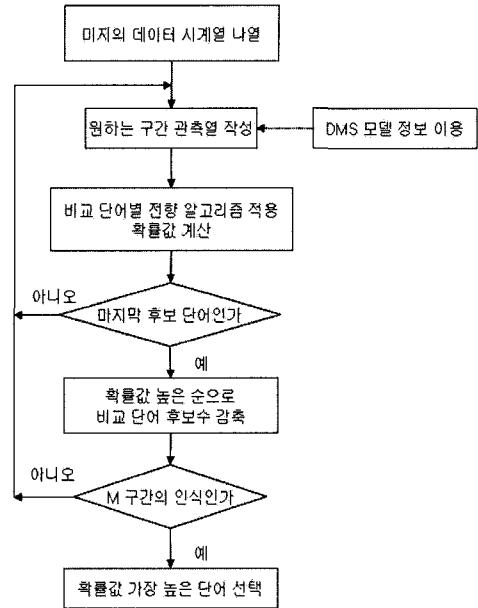


그림 3. 제안된 인식 방법

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.

본 실험에서는 음성을 70Hz ~ 3.5KHz의 대역 통과 필터링을 한 후 8KHz 샘플링과 12bit 양자화를 통해 A/D 변환하였다. 이 데이터는 256 샘플을 한 프레임 단위로 해서 특징을 추출한 후 음성 발생 모델에서 생기는 고주파 감소 현상을 보상하기 위해 프리엠퍼시스 필터 $H(z) = 1 - 0.95z^{-1}$ 를 통과시킨다. 그리고 프레임 추출 시 시작 부분과 끝 부분에 발생하는 왜곡을 보상하기 위해 해밍 창 $W(n) = 0.54 - 0.46\cos(2\pi n/N-1)$ 을 씌운다.

3.1 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 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM

방법 및 다른 여러 방법들의 인식 실험 결과를 표 1에 나타내었다.

여기서, 제안된 DMS 모델과 이중 스펙트럼 특징을 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 | 전체 |
|---------------------------------|-------|-------|-------|
| 제안된 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM | 92.54 | 93.53 | 93.03 |
| DMS 모델을 이용한 HMM | 90.54 | 92.53 | 91.54 |
| FSVQ와 이중 스펙트럼 특징을 이용한 HMM | 90.55 | 92.04 | 91.29 |
| FSVQ를 이용한 HMM | 88.06 | 90.55 | 89.30 |
| 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 | 전체 |
|-------------|-------|-------|-------|
| 2구간으로 나눈 경우 | 92.04 | 93.53 | 92.78 |
| 3구간으로 나눈 경우 | 92.54 | 93.53 | 93.03 |
| 4구간으로 나눈 경우 | 89.55 | 91.04 | 90.30 |

3.2 고찰

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

또한, HMM 인식 방법에서도 일반적인 HMM 보다

MSVQ를 이용한 HMM이 약 6% 정도 인식이 높고, FSVQ를 이용한 HMM이 약 8% 정도 인식이 더 높으며, 마찬가지로 DMS 모델을 이용한 HMM 인식 방법은 약 10% 정도 인식이 높아짐을 알 수 있다. 더구나, 본 연구에서 제안하는 이중 스펙트럼 특징까지를 이용한 방법에서는 약 11.5% 정도까지 인식이 높아진 약 93.03%를 나타냄을 알 수 있다.

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

본 연구에서 제안한 방법은 구간을 의미 있는 특징들을 고려해서 동적으로 잡아주고 이중 스펙트럼 특징에 의해 단시간의 시간정보를 지니게 함으로써 일반적인 HMM 인식 방법보다 약 11.5%의 인식을 증가를 보임을 알 수 있었다.

4. 결론

음성 인식을 하는데 있어서 음성의 특징 파라메타를 잘 나타내고 HMM 모델의 파라메타를 잘 학습시키는 것은 중요한 일이다. 따라서, 본 연구에서는 DMS 모델과 이중 스펙트럼 특징을 이용한 HMM 모델을 제안하고, 이 모델을 이용하여 화자 독립의 단독어 인식 실험을 수행하였고 일반적으로 많이 사용되어 오고 있는 DP 인식 방법과 MSVQ 인식 방법과 DMS 모델을 이용한 인식 실험, 일반적인 HMM 및 MSVQ를 이용한 HMM에 의한 인식 실험 등을 같은 조건하에서 실행하여 비교하였다.

기존의 대표적인 인식 방법인 DP 패턴 매칭 방법의 경우는 기억 용량이 크며, 계산 시간도 많이 걸리는 단점이 있고, MSVQ의 경우는 기억 용량이 적고 인식 시간이 적게 걸리나, 인식이 낮은 단점이 있으며, DMS 모델을 이용한 DMSVQ의 경우는 MSVQ에 비해 인식이 약 7% 정도 높아진다. 또한 일반적인 HMM의 경우도 DP 방법보다는 인식이 높으나 그래도 인식이 낮고, MSVQ를 이용한 HMM의 경우는 인식 시간 및 인식이 일반적인 HMM에 비해서 약 6% 정도 높아지며, FSVQ를 이용한 HMM의 경우는 약 8% 정도 인식이 더 높아지고, DMS 모델을 이용한 HMM에 의한 인식 방법의 경우는 약 10% 정도 더 인식이 높아짐을 알 수 있다.

제안된 인식 실험 방법은 구간을 음소적인 특징에 따라 동적으로 잡아주고 음성의 정적인 특성 뿐만 아니라 동적인 특성까지도 이용함으로 인식이 일반적인 HMM보다 약 11.5% 정도 높은 약 93.03%의 인식을 나타냄을 알 수 있었다.

오인식된 단어들을 살펴 보면, 비슷한 단어들간의 애매한 발음(예를 들면, 오수를 여수에 가깝게, 대천을 대전에 가깝게, 순천을 춘천에 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.

참고문헌

[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, 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, 1990. 9.

[9] 안태옥, "FSVQ에 기초를 둔 HMM을 이용한 음성인식", 한국 컴퓨터 산업교육학회 논문지, 제 2권 제5 호, pp 661-670, 2001. 5.

[10] 안태옥, "FSVQ와 동적 스펙트럼 특징을 이용한 HMM에 기초를 둔 음성인식", 한국 컴퓨터 산업교육학회 논문지, 제 3권 제4호, pp 493-500, 2002. 4.

[11] 안태옥, "FSVQ와 퍼지 개념을 이용한 HMM에 기초를 둔 음성인식", 대한전자공학회, 제40권 sp 편, 제6호, pp 90-97, 2003. 11.

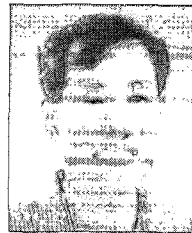
[12] 안태옥, 정의봉, "FSVQ와 개선된 퍼지 개념을 이용한 HMM에 기초를 둔 음성인식", 호원대학교 논문지, 제27집 pp 73-92, 2005. 2.

[13] 안태옥, "구문 분석과 One-Stage DMS/DP를 이용한 연속음 인식", 대한전자공학회, 제 41권 sp 편, 제 3 호, pp 201-208, 2004. 5.

[14] 안태옥, "DMS 모델을 이용한 HMM에 의한 음성인식", 호원대학교 논문지, 제28집 pp 120-132, 2006. 2.

안 태 옥(Tae-Ock Ann)

[정회원]



- 1988년 2월 광운대학교 컴퓨터 공학과 (공학석사)
- 1992년 2월 광운대학교 컴퓨터 공학과 (공학박사)
- 1993년 3월 ~ 현재 : 호원대학교 컴퓨터게임학부 교수

<관심분야>

컴퓨터 구조, 정보통신, 임베디드 시스템, 신호처리 등