DHMM을 이용한 한국어 음성 인식

Korean Speech Recognition using DHMM

안 태 옥* 이 강 성* 유 형 근* 이 형 준* 조 형 제***변 용 규** 김 순 협*

(T. O. Ann, K. S. Lee, H. K. Yoo, H. J. Lee, H. J. Cho, Y. G. Byun, S. H. Kim)

요 약

본 연구는 스펙트럼의 동적 특징을 한 파라메타로 하는 DHMM(Dynamic Hidden Markov Model)을 이용한 단독어 인식에 관한 것으로 정적 스펙트럼 특징뿐 아니라 동적 스펙트럼 특징을 평가할 수 있는 DHMM에 근거한 음성 인식 실험을 논의 한다.

정적특징으로는 LPC cepstrum계수를 이용하였고, 동적특징으로는 LPC cepstrum의 회귀계수를 사용하였다. 이들 두 계수 특징 벡터들을 각각 집단화하여 만든 두 VQ codebook과 입력으로 받아들인 정적 벡터및 동적벡터로 단어들을 DHMM (Dynamic Hidden Markov Model)으로 모델링 하였다.

전체적인 실험에서 기존의 HMM을 이용한 인식실험에서는 88.8%의 인식율을 얻었는데 반해, DHMM을 이용한 인식 실험에서는 92.7%의 인식율을 보였다.

(본 연구는 1989년도부터 2년간 한국 과학 재단의 기초 연구비 지원에 의해 수행된 것임)

ABSTRACT

This paper describes the study on isolated word recognition by using DHMM(Dynamic Hidden Markov Model) which has dynamic feature of spectrum as a parameter. This paper discusses speech recognition experiment based on HMM which can evaluate not only instantaneous spectral features but also dynamic spectral features.

LPC cepstrum parameter is used as a static feature and LPC cepstrum's regression coefficient is used as a dynamic feature. These two feature vectors are quantized by each VQ codebook. DHMM is modeled by receiving static vector and dynamic vector by input.

In the whole experiment, as recognition experiment using DHMM shows 92.7% of recognition rate while the experiment using conventional HMM shows 88.8% of recognition rate, DHMM proved to be a useful model.

I. 서 론

HMM의 기초적인 연구는 1960년대 후반에 Baum 등[1]에 의해서 수행되었다. 그러나, HMM을 이용한 보다 체계적인 연구는 1970년대 중반부터 시작되었으며 IBM Wadson 연구소에서는 문장 입력을 목표로 연속음성 인식과 대어휘 단어 인식에 관한 연구[2][3]를 수행하였고 1980년대에 들어와서는 AT&T Bell 연구소에서 불특정 화자 음성인식에 관한

*광운대학교 전자계산기공학과
**서울산업대학 전자계산학과
***동국대학교 전자계산학과

(본 연구는 1989년도부터 2년간 한국과학재단의 기초연구비지원에 의해 수행된 것임)

인구[4][5]가 대표적으로 수행되었다.

기존의 음성 인식 방법은 크게 패턴 매칭방법과 확률적인 모델에 의한 방법으로 나눌 수 있다. 패턴 매칭 방법은 각 화자를 대표할 수 있는 패턴들을 미리 작성한 다음, 시험패턴과 표준패턴의 유사도를 측정하여 비교해 가장 근사한 것으로 인식되는 것으로 생각하는 방법이다.

음성의 특징은 개인의 차가 클 뿐만 아니라 같은 사람이 같은 내용을 말 하려고 해도 불안정성이 있다. 확률적인 모델링은 패턴매칭을 특수한 경우로서 포함하는 좀더 일반화된 방법으로서 실제 음성의 불완전성을 보다 정확하게 반영하며 확률적인 개념을 이용하여 이론적 전개가 쉽고 언어처리등을 통합하기가 쉽다.

음성의 특징은 스펙트럼의 변화에 많은 정보가 포함되어 있으므로 스펙트럼의 동적 변화도를 고려해야 한다는 것은 잘알려진 사실이지만 기존의 HMM 은 정적인 스펙트럼의 불완정성이나 시간의 흐름을 확률적으로 모델링하고 있을뿐 실제로 동적 특징을 파라메타로 고려하고 있지 않다. 따라서 본 연구에서는 기존의 정적인 특징뿐만 아니라 동적인 특징까지도 고려하는 HMM을 이용하여 음성 인식 실험을 하였다.

동적인 특징을 고려한 연구로 시간의 미분 정보[7]같은 이웃한 spectral 특징의 연결을 Paritz 등이 사용하였지만 스펙트럼이 시간과 함께 어떻게 변화하는지에 관한 정보가 없다. 몇개의 프레임상에 이웃한 스펙트럼 특징을 복잡하지만 HMM으로 모델링[8]할 수 있으나 긴 구간상의 시간변화에 따른 패턴을 모델링하기 위해서는 많은수의 파라메타가 필요하고 파라메타를 추정하는데 많은 학습 데이타가 있어야 한다. 또한 HMM을 바탕으로 한 인식 시스템에 순간 스펙트럼과 스펙트럼의 시간열의 1차 미분계수를 사용하여 각 특징[3]을 독립적으로 VQ 하는 이 인식 기술은 HMM을 나타내는데 효과적이지만 기존의 HMM과 비교하여 많은 파라메타가 필요하다.

따라서, 본 연구에서는 정적 spectral특징과 동적 spectral특징의 상관관계가 적다[6]는 연구 결과를 바탕으로 모델의 각 상태에서 이들의 특징을 독립적으로 계산함으로써 시간 변화에 따른 패턴을 모델링하고 인식[5]

Ⅱ. 음성 신호의 분석

본 연구에서는 시작점과 끝점을 검출하는데 ZCR (Zero crossing rate)와 LE(Log Energy)을 이용한 Rabiner및 Samber의 끝점검출방법[1]을 사용하였고, 음성파형은 3.5KHz의 차단 주파수를 갖는 저역 여파기를 통과하여 8KHz로 샘플링하였으며, 각 샘플들은 12비트로 구성하였다. 16ms를 한프레임으로 16ms씩 이동하면서 10차로 LPC cepstrum계수를 구했다. 이 때 Pre-emphasis은 전달함수

$$H(z)=1-az^{-1} \quad (a=0.95) \tag{1}$$

을 사용하였다.

또한 특징 파라메타는 정적인 특징으로 LPC cepstrum 계수를 사용하였고, 동적인 특징으로 이 cepstrum의 회귀 계수를 사용하였다.

1. 켑스트럼 파라메타 추출

음성 인식을 사용하는 정적 파라메타에는 LPC, 구조만[ ], LSP등 여러가지가 있지만 이들 파라메타를 상호 비교한 Nakagawa와 Shikano등의 실험결과[12][13], LPC 파라메타보다 LPC cepstrum의 파라메타가 음성을 인식 하는데 좋다는 고찰에 따라 본 실험에서는 LPC cepstrum을 특징파라메타로 사용하였다.

2. Dynamic 파라메타추출

대부분 음성 인식 시스템은 정적인 파라메타의 계수만을 사용한다. 근래들어 연구자들은 스펙트럼내에서 능동적인 변화를 측정하는 계수를 사용하기 시작 했다. Paul[18]등은 LPC cepstrum 계수를 사용하였고, Furui[19]은 스펙트럼 내에서 변화를 측정하기위해 선형 회귀 계수를 사용하였다. 음성의 특징 파라메타를 분석할때 스펙트럼에 있어서 일시적인 변화를 나타내는데 이용하는 동적 특징이 중요한 역할을 한다. 근본적으로 스펙트럼의 동적 특징이

나 할 수 있는 기울기를 측정하는 회귀 계수는 다음과 같다.

$$D_i(t)=\frac{\sum_{n=-k}^{k} n\cdot\{C_i(t+n)\}}{\sum_{n=-k}^{k} n^2} \qquad (2)$$

여기서 $C_i(t)$는 발음의 t번째 프레이미의 i번째 계수이고, D는 회귀계수이다. 기울기는 −k부터 k까지 측정한다. 이하부터는 이를 회귀 계수라고 한다. 정적 스펙트럼과 동적 스펙트럼을 아래 그림 1에 보여준다.

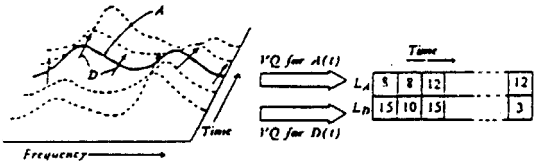

그림 1. 정적스펙트럼과 동적 스펙트럼
Fig. 1. Instantaneous spectra and dynamic spectra

Ⅲ. 코드북 작성

코드북을 구성하는 방법에는 단어마다 각각의 코드북을 두는 방법과 전체 단어를 하나의 코드북으로 두는 방법이 있다. 그런데 전자 방법은 단어가 증가 함에 따라 코드북 크기가 증가하여 검색 시간이 크게 증가한다. 후자의 방법은 시간이 적게 걸리며 음성 인식 성능이 전자의 방법에 비해 별차이가 없으므로 전체적으로 코드북을 작성했다.

VQ를 이용한 음성 인식 시스템에서 코드북은 학습용 데이타의 특성이 잘 나타나도록 코드북을 만들어야 한다. K-means 알고리즘을 이용하여 LPC cepstrum 계수와 회귀 계수를 바탕으로하여 각각의 codebook을 만들었다.

K-means 알고리즘은 Cluster center수(k), 초기 cluster 값, 샘플들이 정해지는 순서, 데이타의 기하학적 성질등에 의해 영향을 받아 실행될 때 마다 동일한 결과가 얻어지지 않으나 계산이 간편하며 다른 방법에 비해 성능이 떨어지지 않는다. 따라서 이 K-means 알고리즘을 이용하여 코드북을 작성하였다

이 때, 사용되는 두 벡터간의 거리는 다음과 같이 구했다.

$$d(x,y)=W(C_{x0}-C_{y0})^2+\sum_{i=1}^{P}(C_{xi}-C_{yi})^2 \qquad (3)$$

여기서, x와 y는 두개의 벡터를 나타내고 W는 가중치이며, C_i는 특징 파라메타를 나타낸다. P는 10으로, W는 0.04로 설정하였다.

정적 코드북의 크기는 128로 하였으며, 회귀 계수를 구할때 k=2로 하여 식(2)로부터 기울기를 구하였으며 동적 코드북의 크기도 128로 하여 clustering 하였다. 이 각각의 코드북은 DHMM을 모델링 할때 관측 심볼의 수로 사용된다.

Ⅳ. HMM에 의한 인식

HMM은 천이들에 의해 서로 연결된 상태들의 모임으로서 각 천이에는 2가지 종류의 확률이 관련되어 있다.

하나는 현재의 천이가 이루어질 천이 확률이고, 또 하나는 천이가 이루어졌을 때 유한개의 관측 대상으로부터 각 출력 심벌이 출현될 조건부 확률을 규정하는 출력 확률 밀도 함수(pdf)이다. 그림 2는 2개의 상태와 2개의 출력 심벌 A, B를 가진 HMM의 예이다.

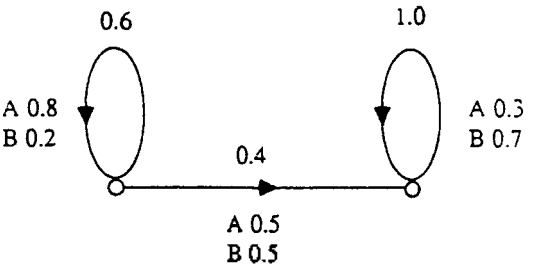

그림 2. 간단한 HMM의 예
Fig. 2. An example of a simple HMM

기호들은 다음과 같이 정의한다.

(기 호)

상태수 : N

관측 심볼수 : M

상태 집합 : $Q=\{q_1, q_2, \cdots, q_N\}$

심볼 집합 : $V=\{v_1, v_2, \cdots, v_M\}$

상태열의 크기 : $t=1, 2, \cdots, T$

t번째 관측 심볼열이 상태 q_i에 있고, t+1 번째 관측 심볼열이 상태 q_j를 선택할 확률

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

t번째 관측 심볼열이 q_j 상태에서 심볼 v_k를 선택할 확률

$$B=\{b_j(k)\},\ b_j(k)=pr(v_k \text{ at } t \mid q_j \text{ at } t) \quad (5)$$

초기 상태에서 상태 q_i를 선택할 확률

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

관측열 $O=O_1, O_2, \cdots, O_T$

이상의 정의를 이용하면, HMM은 모델 $\lambda=(A,B,\pi)$로 표시 할 수 있는데 이 모델을 실제 응용하는데는 세가지 해결해야할 문제점이 있는데 다음과 같이 해결한다.

1) 관측열 $O=O_1, O_2, \cdots, O_T$와 모델이 주어졌을 때 관측열이 나올 확률 $P_r(O|\lambda)$를 계산하는 방법은 forward-backward 알고리즘을 이용하여 해결한다.
2) 관측열 $O=O_1,O_2,\cdots, O_T$와 모델이 주어졌을 때 최적의 상태열 $I=i_1, i_2,\cdots, i_T$를 구하는 문제는 Viterbi 알고리즘을 이용하여 해결한다.
3) $P_r(O|\lambda)$를 최대화 하기 위해 λ를 조정하는 문제는 Baum-Welch의 재추정 알고리즘을 이용하여 해결한다.

위와 3가지 방법에서 알 수 있는 바와 같이 Baum-Welch 의 재추정 알고리즘에 의해 HMM을 학습시키며, Viterbi 알고리즘이나 forward 알고리즘을 이용하여 인식 실험을 행한다.

V. DHMM에 의한 인식

DHMM은 정적 스펙트럼의 특징 파라메타와 동적 스펙트럼의 특징 파라메타를 함께 모델링한것이다. 그림 3은 정규화된 log power(P), 그것의 동적특징(PD), 정적 특징(A), 동적 특징(D) 사이의 상관관계의 절대치를 보여준다.

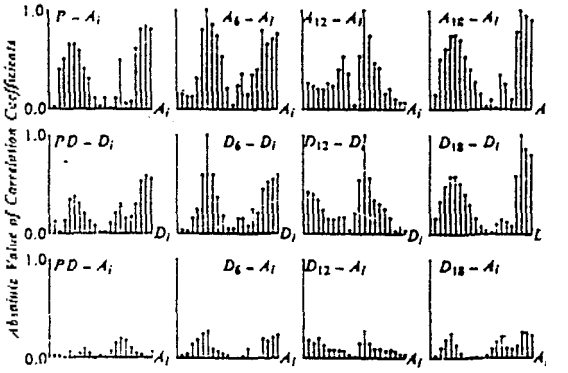

그림 3. P, PD, A와 D특징 사이의 상관관계
Fig. 3. Correlation between P, PD, A and D

즉 A와 D사이의 상관 관계가 A와 A나 D와 D의 상관관계 보다도 매우 작다[7]는 사실을 이용하여 동적 스펙트럼의 특징을 비탕으로 한 DHMM은 HMM에 비해 거다른 계산의 증가 없이 모델링 할 수 있다. 이러한 모델을 구성하는 요소를 정의해보자.

여기서 상태 S와 t번째 관측열 O_t는 다음과 같다.

$P(O_t,S)=b_S(O_t)$라 하자.

$$S=(M_t, M_D) \quad (7)$$

$$O=(O_{tt}, O_{Dt}) \quad (8)$$

$$P(O_t,S)=P(O_{tt}, O_{Dt}|S) \quad (9)$$

정적특징 (O_{tt})과 동적특징(O_{Dt})은 서로 상관관계가 거의 없으므로 식(10)과 같이 쓸 수 있다.

$$P(O_t|S) \equiv P(O_{tt}|S)\ P(O_{Dt}|S)$$

$$=P(O_{It}, M_I, M_D)\ P(O_{Dt}, M_I, M_D)$$
$$=P(O_I, M_I)\ P(O_D, M_D) \quad (10)$$

여기서

$$b_S(O_t)=P(O_t|S)$$
$$\equiv P(O_{It}, M_I)\ P(O_{Dt}|M_D)$$
$$=b_s(O_{It})\ b_s(O_{Dt}) \quad (11)$$

로 된다.

여기서, $bs^I(O_{It})$은 상태 S에서 정적벡터 O_{It}가 나올 확률이며 $b_s^D(O_{Dt})$는 상태 S에서 동적벡터 O_{Dt}가 나올 확률이다.

1. DHMM 모델링

DHMM에서 초기 파라메타들로 부터 $P_r(O|\lambda)$를 최대로 하는 $\lambda=(A,B,\pi)$를 재추정하는 Baum-Welch reestimation 알고리즘은 다음과 같다.

$$\tilde{a}_{ij}=\frac{\sum_{t=1}^{T-1}\bar{\alpha}_t(i)a_{ij}b_j(O_{t+1})\bar{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1}\bar{\alpha}_t(i)\bar{\beta}_t(i)} \quad (12)$$

$$\tilde{b}_j^I(k)=\frac{\sum_{t=O_{It}=V_{Ik}}\bar{\alpha}_t(j)\bar{\beta}_t(j)}{\sum_{t=1}^{T}\bar{\alpha}_t(j)\bar{\beta}_t(j)} \quad (13)$$

$$\tilde{b}_j^D(k)=\frac{\sum_{t=O_{Dt}=V_{Dk}}\bar{\alpha}_t(j)\bar{\beta}_t(j)}{\sum_{t=1}^{T}\bar{\alpha}_t(j)\bar{\beta}_t(j)} \quad (14)$$

위의 식을 이용하여 정적인 특성뿐만 아니라 동적 특성을 포함하는 DHMM을 모델링 한다.

2. DHMM을 이용한 인식

DHMM의 경우는 forward 알고리즘을 이용하여 인식하였다. 왜냐하면, 기존의 HMM 인식으로 forward 알고리즘과 Viterbi 알고리즘을 이용한 인식의 결과를 비교해 본 결과 forward 알고리즘을 이용한 인식 방법이 더 좋은 인식률을 보였기 때문이다. 따라서, 본 절에서는 forward 알고리즘에 대해서만 설명하겠다.

Forward variable $\alpha_t(i)$은 식 (15) 처럼 정의된다.

$$\alpha_t(i)=P_r(O_{It}, O_{Dt}) \cdots\cdots (O_{It}, O_{Dt}), i_t=q_i|\lambda) \quad (15)$$

다음과 같은 recursion에 의해서 $P_r(O|\lambda)$을 구할 수 있다.

step 1. 초기화

$$\alpha_t(i)=\pi_i b_i(O_{I1})b_i(O_{I1}),\ 1\leqq i\leqq N \quad (16)$$

step 2. t=1, 2,… T-1, $1\leqq i\leqq N$

$$\alpha_{t+1}(j)=[\sum_{i=1}\alpha_T(i)a_{ij}]*b_j(O_{It+1})b_j(O_{Dt+1}) \quad (17)$$

step 3. 그러면

$$P(O|\lambda)=\sum_{i=1}^{N}\alpha_T(i) \quad (18)$$

따라서, 최종 관측열까지의 확률은 step 3에 의해 구해진다.

3. DHMM의 구현

HMM을 실제 모델링하는데는 세가지 문제점이 대두되는 데 마찬가지로 DHMM을 구현하는데도 세가지 문제점이 있다. 첫째로 DHMM의 파라메타를 어떻게 초기화할 것인가라는 문제이고, 둘째는 T가 증가함에 따라 이들의 값이 지수함수적으로 감소하여 곧 underflow를 발생시킨다는 점이며, 세째로 학습 데이타의 양이 충분히 많지 않을 경우 학습과정에서 어떤 심볼이 모델의 어느 상태에는 나타나지 않는것으로 추정되었더라도 실험 과정 중에서 이 심볼이 나타나는 경우가 있다는 점이다. 본 연구에서

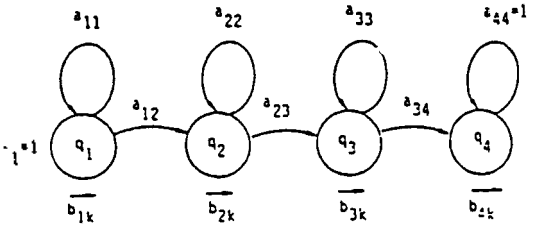

그림 4. HMM의 모델
Fig. 4. Hidden Markov model.

는 모델을 ... 4개 상태의 left to right 모델을 사용하였다.

1) 초기화

HMM의 초기화를 위한 제한조건을 다음과 같다.

$$\pi_1=1,\ \pi_i=0,\ i\neq 1 \quad (19)$$

$$a_{ij}=0,\ \text{for } j<i \text{ and } j\geq i+2 \quad (20)$$

$$\sum_{j=1}^{N} a_{ij}=1,\ i=1,2,\cdots N \quad (21)$$

$$\sum_{k=1}^{M} b_j(k)=1,\ j=1,2,\cdots,N \quad (22)$$

위의 식들을 다음과 같이 초기값을 설정하여 위의 조건을 만족하도록 식(21), (22)의 파라메타를 재계산하였다.

$$a_{ii}=\frac{1}{2}+\varepsilon,\ a_{ij}=1-a_{ii},\ (\varepsilon>0) \quad (23)$$

$$b_j(k)=\frac{1}{M}+\varepsilon,\ (\varepsilon>0) \quad (24)$$

또는,

$$b_j^I=\frac{1}{M}+\varepsilon,\ b_j^D=\frac{1}{M}+\varepsilon,\ (\varepsilon>0) \quad (25)$$

이때 ε은 작은 랜덤 실수이다.

2) Scaling 법

실제로 forward-backward 알고리즘을 계산할 때는 T가 증가 함에 따라 이들의 값이 지수함수적으로 감소하여 곧 underflow를 발생시킨다. 변수의 값들이 컴퓨터의 계산 영역에 있도록 i에 무관한 scaling 값을 $\alpha_t(i)$와 $\beta_t(i)$에 각각 곱하여 이를 해결한다.

scale 요소, C_t는

$$C_t=[\sum_{j=1}^{N}\alpha_t(j)]^{-1} \quad (26)$$

이다

$$\sum_{i} C_t\alpha_t(i)=1,\ 1\leq t\leq T \quad (27)$$

가 된다. 식 (12), 식 (13)와 식 (14)의 scale된 forword, backward 확률식은

$$\tilde{a}_{ij}=\frac{\sum_{t=1}^{T-1} C_t\alpha_t(i)a_{ij}b_j^I(0_{It+1})b_j^P(0_{Dt+1})\overline{\beta}_{t+1}(j)D_{t+1}}{\sum_{t=1}^{T-1} C_t\alpha_t(i)\ \beta_t(i)D_{t+1}} \quad (28)$$

$$\tilde{b}_j(k)=\frac{\sum_{t=1,\ldots}^{T-1} C_t\alpha_t(i)a_{ij}b_j^I(0_{It+1})b_j^P(0_{Dt+1})\ \overline{\beta}_{t+1}(j)D_{t+1}}{\sum_{t=1}^{T} C_t\alpha_t(i)\ \overline{\beta}_t(i)\ D_t} \quad (29)$$

가 된다.

여기서

$$C_t=\sum_{t} c_t,\ D_t=\sum c_t$$

이다.

3) Smoothing

학습 데이타의 양이 충분히 많지 않은 경우 학습과정에서 어떤 심볼이 모델의 어느 상태에서는 나타나지 않는것으로 추정되었더라도 실험과정중에서 이 심볼이 나타나는 경우가 있다. 이때 모델이 추정한 관측 확률이 0이므로 전체 음성 발생 확률이 0이 되어 확률 계산을 제대로 할 수 없다. 그러므로 이들 확률을 smoothing함으로, 훈련되지 않은 심볼이라도 나타날 가능성을 고려하여 계산하는 것이 필요하다. 즉 최소치는 ε>0을 설정한다.

즉, 식 (29)을 아래와 같이 조정한다.

$$b_j(k)=(1-h)\frac{b_j(k)}{\sum_{l=1}^{N-1} b_j(l)},$$

$$\text{if } b_j(k)\geq\varepsilon \quad (30)$$

$$b_j(k)=\varepsilon,\ \text{otherwise}$$

여기서 $L_2(k)$ ε인 샘플의 개수이다. 위와 같이 계산을 중단하는 것이 실험적으로 우수하다고 알려진 ε의 값은 $10^{-10} \le \varepsilon \le 10^{-3}$의 넓은 범위에서 거의 같은 성능을 갖는다.[13] 본 연구에서 설정한 값은 0.00001이다.

Ⅵ. 실험 및 고찰

1. 실험 조건 및 음성 데이타

본 연구의 음성 데이타는 방음 장치가 제대로 되지 않아 잡음이 있는 방에서 다이나믹 마이크를 통해 얻은 음성파형을 3.5 KHz의 차단 주파수를 갖는 저역 여파기를 통과하여 8KHz로 샘플링하였다. 각 샘플들은 12비트로 구성되었으며, 16ms을 한프레임으로 16ms씩 이동하면서 특징 벡터를 구했다. 음성의 각 프레임은 Hamming window를 사용하여 처리하였다. 특징 벡터로는 10차의 cepstrum 계수를 사용하였고 이를 바탕으로 dynamic 계수를 추출하였다. 그리고 이것들은 K-means 알고리즘에 의해 정적 특성과 동적 특성의 코드북이 작성되었다. 본 연구의 실험을 위해 발음한 화자는 남학생 5명, 여학생 1명, 성우 1명, 그리고 아나운서 3명으로 총 10명이다. 또한, 인식 대상어는 전국 146개 DDD 지역명으로 각 화자가 5번씩 발음하였다.

2. 인식 실험

본 연구에서는 기존의 HMM과 DHMM과 비교하여 실험하였다. 이 때 기존의 HMM의 경우 인식시에 forward 알고리즘에 의한 인식률과 Viterbi 알고리즘에 의한 인식률을 비교하였고, 인식 실험에서 forward 알고리즘에 의한 인식률이 더 높은 관계로, DHMM에 의한 인식실험은 forward 알고리즘만 이용하였다.

전체적인 DHMM 음성 인식 시스템의 개략도는 그림 5와 같다.

2. 기존의 HMM 인식 실험

이 실험에서 이용된 정적 코드북은 10명의 화자가 각각 2번씩 발음한 것 중에서 128개의 코드워드를 선정하였으며, 모델을 만드는데 있어서 재추정 알고

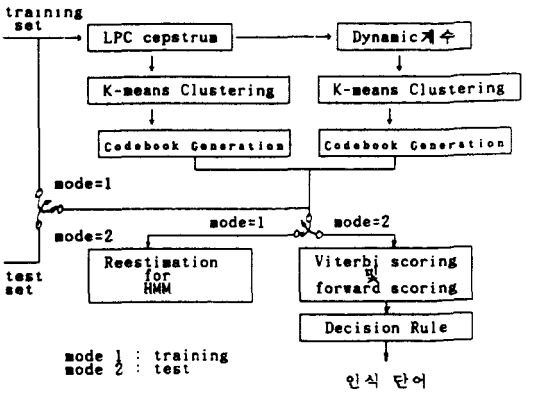

그림 5. 전체적인 DHMM의 음성 인식 시스템의 개략도.
Fig. 5. Overall block Diagram of DHMM using dynamic feature.

리즘에 사용된 학습용 데이타는 10명의 화자가 각각 3번씩 발음한 것으로 수행하였다.

그리고, 각 단어에 대한 인식은 forward 알고리즘과 Viterbi 알고리즘으로 수행하였으며, 각 알고리즘에 의한 인식 실험의 결과는 다음 표 1과 같다.

표 1. 기존의 HMM의 인식률

화 자	Viterbi 알고리즘에 의한 인 식		forward 알고리즘에 의한 인 식	
	인식률(%)	에러갯수(개)	인식률(%)	에러갯수(개)
A	85.3	107	86.6	98
B	82.1	131	81.8	133
C	91.9	59	91.9	59
D	89.5	77	90.1	72
E	86.8	96	88.8	82
F	89.5	77	89.7	75
G	88.2	86	89.3	78
H	92.9	52	92.9	52
I	82.2	130	82.6	127
J	93.2	50	93.8	45
총 계	88.2	865	88.8	821

2.2 DHMM 인식 실험

DHMM은 정적 코드북과 동적 코드북이 동시에 필요한데 본 연구에서는 HMM에서 사용한 정적 코드북을 그대로 이용하였으며, 또한 같은 데이타로 코드북의 크기 128의 동적 코드북을 작성하였다.

DHMM 모델 작성시에도 HMM의 경우와 마찬가

자는 10명의 화자가 각각 3번씩 발음한 것으로 수행하였다.

[illegible] HMM의 인식 결과에 따라 forward 알고리즘으로 인식실험을 수행하였다. 인식 실험 결과는 표 2에 나타내었다.

표 2. DHMM에 의한 인식률

화자	Forward 알고리즘에 의한 인식	
	인 식 률(%)	에 러 개 수
A	91.918	59 개
B	85.753	104 개
C	93.836	15 개
D	93.151	50 개
E	92.603	54 개
F	94.932	37 개
G	94.521	40 개
H	95.205	35 개
I	88.767	82 개
J	96.164	28 개
총 계	92.687	534 개

3. 고찰

본 실험은 DDD지역명을 인식 대상어로 삼았으며, HMM 및 DHMM을 학습하는데 있어서 10명이 146개 DDD지역명을 5번 발음한 음성 데이타중에서 10명 모두 3번씩 발음한 것이 모델링하는데 사용되었다.

표 1에서 볼 수 있는 바와 같이 기존의 HMM으로 모델링하여 실험한것은 인식 알고리즘으로 forward 알고리즘과 Viterbi 알고리즘으로 나누어 실험하였는데, 실험 결과 forward 알고리즘에 의한 인식 실험의 경우는 88.8%의 인식률을 나타내었고, Viterbi 알고리즘에 의한 인식실험의 경우는 88.2%의 인식률을 나타내었다. 이 실험에서 화자 B(남성 성우)만이 Viterbi 알고리즘에서 인식률이 더 좋았다. 따라서, Viterbi 알고리즘보다 forward 알고리즘에 의한 인식이 더 좋음을 알 수 있었다.

또한, 표 2는 DHMM으로 모델링하여 인식 실험한 결과를 나타내었다. 이 DHMM의 경우는 HMM 인식의 결과로 forward 알고리즘으로 인식 실험하였다. 실험 결과 92.7%의 인식률을 얻었다.

[illegible] 비교에서 알 수 있는 바와 같이 정적 특성만을 이용하는 HMM에 비하여 동적 특징까지 이용하는 DHMM이 그다지 인식 시간이 증가 없이 인식률이 향상됨을 알 수 있다.

또한, 각 개인별 인식률을 살펴 보면, 화자 B가 인식률이 가장 나쁘다. 그 이유는 DDD 지역명을 읽을 때, 앞 단어와 뒷 단어 사이에 시간간격을 두지 않고 발음한 관계로 단어를 샘플링하는 과정에서 앞 단어와 뒷 단어가 많은 영향을 주었을 뿐 아니라 어떤 것은 다음 음성이 섞여 샘플링 된 것도 있기 때문이다. 화자 I(남학생)의 경우는 라디오 잡음이 발음한 음성과 거의 같은 크기를 가진 관계로 인식률이 좋지 않은 것 같다. 그외에 아나운서(화자 E, F, G)와 남학생 C, D, J은 평균 인식률 이상의 인식률을 나타내었으며, 여학생 H는 대부분의 나머지 사람보다 상당히 높은 95.2%의 인식률을 나타내었다. 여기에서 여성 화자가 인식률이 높은 이유는 발음이 비교적 분산 값이 적었기 때문이다.

전체적으로 볼 때, 기존의 정적 특성 벡터만을 이용하는 HMM 모델링 방법은 동적인 스펙트럼의 변화성을 갖지 않으므로 일시적인 변화에 보다 적절히 대응하지 못하는 것으로 생각된다. 그런데, DHMM 의 경우는 정적 특징 벡터에 이 동적 특징까지도 고려해 줌으로 이러한 것들을 보완할 수 있는 것으로 생각된다.

Ⅵ. 결 론

본 연구에서는 기존의 정적 특징만을 바탕으로한 HMM과 동적 특징까지도 고려한 DHMM과 비교 연구하였다. 또한 인식 알고리즘으로 Viterbi 알고리즘과 forward 알고리즘을 비교하였다.

실험 결과, HMM의 경우에 Viterbi 알고리즘에 의한 인식률은 88.2%이고, forward 알고리즘에 의한 인식률은 88.8%로 forward 알고리즘에 의한 인식률이 더 좋음을 알 수 있었다. 이를 바탕으로, DHMM에서 forward 알고리즘을 이용하여 인식하였을 경우 92.7%의 인식률을 얻는다. 따라서, DHMM에 의한 인식이 기존의 HMM에 의한 인식보다 인식률이

상당히 향상시킬 수 있다.

그 이유는 기존의 HMM에 의한 인식 방법은 정적 특징 파라메타만을 이용하므로 음성에 있어서의 동적인 특징을 반영하지 못함을 알 수 있으며, 반면에 정적특징 뿐만 아니라 동적특징 파라메타를 이용하는 DHMM의 경우는 그 실험결과가 보여주듯이 화자독립 인식의 인식률 성능을 향상 시켰다. 이것으로 정적특징만을 사용한 것보다는 동적특징을 함께 사용하는 것이 훨씬 더 음성의 특징을 잘 나타낸다는 것을 알 수 있다.

앞으로, 정적 특징 뿐만아니라 동적 특징을 사용한 DHMM의 모델을 사용하여 음소나 연속음 인식에 적용하여 좋은결과를 얻을 수 있을것으로 생각된다.

참 고 문 헌

1. L.E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technicque occurring in the statistical analysis of probabilistic function of Markov chains," Ann. Math. stat. 41, 164-171(1970)
2. F. Jelinek, "The development of an experimental discrete dictation recognition," Proc. IEEE 73, 1616-1624 (1985)
3. L.R. Bahl, P.F. Brown, P.V. de Souza and R.L. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," Proc. ICASSP 86, 49-52(1986)
4. L.R. Rabiner, S.E. Levinson and M.M. Sondhi, "On the application of vestor quantzation and hidden markov model to speaker-independent isolated word recognition," Bell Syst. Tech. J. 62, 1075-1105 (1983).
5. B. H.Juang, L.R. Rabiner, S.E. Levinson and M.M. Sondhi, "Recent development in the application of hidden Markov model to speaker-independent isolated word recognition," Proc. ICASSP 85, 9-12(1985).
6. A. B. Poritz and A.G. Richer, "On the hidden Markov models in isolated word recognition," Proc. ICASSP' 86, 14-3, April 1987.
7. M. Nishmura, "HMM-based Speech Recognition Using Dynamic spectral Feature," Proc. ICASSP'89 298-301, May 1989.
8. C.J. Wellekens, "Explicit time correration in hidden Markov models in isolated word recognition," Prdc. ICASSP'87, 10-7, April 1987.
9. Seiich Nakagawa, Mitsunori Sakamoto, "Evalution of FFT cepstrum and LPC cepstrum for speech spesker Recognition," 일본 전자 통신 학회 논문집, vol. J66-A No.12, pp 1199-1206 1983.
10. Seikano, k. Kohda, M. "On the LPC distance Measures for Vowel Recognition in continuous utterances," IEEE Japen, D. J63-D157, May. 1980.
11. Paul, D.B., Lippmann, R,P., Y., Weinstein, C. Robust HMM-Based Techniques for Recognition of Speech Produced under Stress and in Noise. In Speech Tech. April, 1986.
12. Furui, S. "Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum," IEEE trans. on Acoustics, Speech, and Signal Processing ASSP-34(1):52=59, Feb. 1986.
13. L.R. Rabiner, B.H. Juang "An introduction to Hidden Morkov Models," IEEE ASSP Magazine pp. 4-16 Jan. 1986.

▲안태욱 : 8권 4호 참조
▲이강성 : 8권 3호 참조
▲유형곤 : 8권 3호 참조
▲이형준 : 6권 4호 참조
▲조형제 : 8권 5호 참조 조영제
▲변용규 : 8권 5호 참조
▲김순협 : 8권 5호 참조