

의 발생빈도가 일정하지 않고 각 음절당 학습샘플의 수도 일정하지 않다. 충분한 학습 샘플을 확보하기 위해 화자 3명의 음성 데이터를 혼합하여 학습하였으며 다수화자 상태로 학습과 인식을 실시하였다.

표 2. 4연 숫자음

4연 숫자음						
2087	5732	9601	4156	1199	1398	6843
0712	5267	6633	2409	7954	1823	6378
8877	3510	8065	2934	7489	2244	4621
9176	3045	8590	5500	6972	5861	3649
0316	7083	8194	9205	1427	2538	4750
단독 숫자음						
1(일) 2(이) 3(삼) 4(사) 5(오) 6(육) 7(칠) 8(팔) 9(구) 0(공)						

표 3. 신문 사설에서 발췌된 연속음성

- 1) 우리의 생활문화가 문화국민의 품격을 잃고 있다.
- 2) 반이란 그 나라 그 사회의 문화의 척도다.
- 3) 말이 잘 정리되고 품격을 유지하는 사회
- 4) 우리는 세계적으로 손색없는 우수한 말과 글을 가지고 있다.
- 5) 가정에서의 말의 교육이 전무한 실정이다.
- 6) 학교나 사회에서의 언어교육이 중요하다.
- 7) 가정에서의 언어교육이 사회교육으로 연결된다.
- 8) 우리사회의 언어 현실이라 해도 과언이 아니다.
- 9) 우리의 국력과 문화수준에 맞는 언어 생활의 정화가 시급하다.
- 10) 방송이나 매스컴 종사자들의 엄격한 언어 통제가 필요하다.

4.2 모델의 구조와 학습

본 실험에서 구성한 연속음성 인식 시스템은 HMM 모델의 구조에 따라 연속분포 HMM과 지속시간제어 연속분포 HMM, 2종류이며 인식 알고리즘은 O(n)DP법이다. 또한 인식의 단위는 음절과 단어 단위가 혼합된 형태이며 학습은 모두 음절 단위로 실시된다.

그림 5에 음절 HMM 모델의 구조와 음절 HMM을 연결하여 단어 단위 HMM을 작성하는 방법을 나타냈다. 단어 단위 인식의 경우 이미 학습된 음절단위 HMM 모델이 앞 음절의 최종상태와 후속 음절의 초기 상태가 널(null) 천이로 연결되도록 하는 연결기와 단어사전을 작성하여 자동 구성되도록 하여 모델의 작성에 필요한 학습 시간과 노력을 대폭 줄일 수 있다.

각 음절 HMM 모델은 모두 5상태 4출력 분포 HMM이며 출력확률은 연속출력 확률분포 HMM을 이용했다.

한편 이산 지속시간제어 HMM의 경우 천이 확률은 모두 1로 설정되기 때문에 그림 6에서 자기 루프 천이는 제거되며 지속시간 확률분포의 초기치는 최대 허용 시간의 범위 안에서 균일 분포로 준다. 여기서 최대 허용시간은 학습에 이용된 음절 가운데 가장 길게 발생된 음절은 118 프레임이므로 한 음절의 최대 허용 지속시간은 120ms로 하였으며 각 상태당 30프레임으로 할당되어야 하나 초성 지음은 모음의

발생시간에 비해 매우 짧으므로 초성 지음과 모음의 결합 부분에는 2개의 상태가 할당될 확률을 높이기 위하여 그림 6과 같이 상태 1과 상태 2의 최대 허용 지속시간을 1/2로 설정하였다.

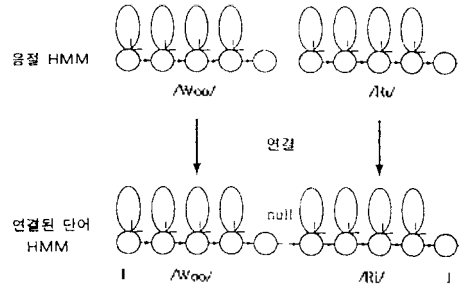


그림 5. 음절 HMM과 단어 HMM의 구조

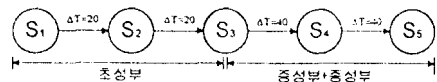


그림 6. 이산 지속시간제어 확률분포 HMM 모델의 구조와 최대 허용시간

4.3 연속음성 인식 결과 및 고찰

본 실험에서 구성한 연속음성 인식 시스템은 HMM 모델의 구조에 따라 연속분포 HMM과 지속시간제어 연속분포 HMM, 2종류이며 인식 알고리즘은 O(n)DP법이다. 또한 인식의 단위는 음절과 단어 단위가 혼합된 형태이며 학습은 모두 음절 단위로 실시하였다.

4.3.1 실험1

표 4에는 연속분포 HMM 모델과 이산 지속시간 제어 확률분포 HMM에 대한 숫자음의 인식 결과를 나타내었다. 인식결과 연속분포 HMM의 경우 평균 80.7%의 인식 결과를 나타내었으며 지속시간을 첨가한 경우는, 평균 93.1%의 인식 결과로 약 12.4%의 향상된 결과를 나타내었다. 이것은 기본적인 연속분포 HMM의 경우 자기상태에 지속될 확률(자기 루프 천이확률)이 지수 함수적으로 감소하여 각 단어나 음절의 앞뒤에 15 프레임 미만의 짧은 음절이 인식되는 삽입에 의한 오인식이 많은 반면 지속시간 제어의 경우가 음절의 지속시간 확률에 의하여 삽입에 의한 오인식을 대폭 감소시키기 때문인 것으로 분석된다.

4.3.2 실험2

신문사설에서 발췌한 연속음성을 CHMM으로 학습하여 화자종속 및 독립에 대해 연속 음성 인식 실험하였다. 표 5의 결과에서 종속은 82.1%, 독립에는 54.2%인식률을 얻었다.

4.3.3 실험3

표 6에는 연속분포 HMM 모델과 이산 지속시간 제어 확률분포 HMM에 대한 신문사설의 인식 결과를 나타내었다. 인식결과 연속분포 HMM의 경우 평균 54.2%의 인식 결과를 나타내었으며 지속시간을 첨가한 경우는 평균 68.9%의 인식 결과로 약 14.2%의 향상된 결과를 나타내었다.

*인식률=(전음절수-삽입-탈락-치환)/전음절수*100

표 4. CHMM과 DDCHMM에 의한 숫자음인식결과.

CHMM Word Spotting기법을 이용한 연속음성 인식에 관한 연구

김수훈 이종진** 허강인*

* 동아대학교 전자공학과

** 장안전문대학 전자계산학과

A Study on the Recognition of Continuous Speech using
CHMM Word Spotting

Soo-Hoon Kim Jong-Jin Lee** Kang-In Hur*

* Dept. of Electronic Eng., Dong-A Univ.

** Dept. of Computer Science, Jang-Ahn Junior College.

요약

본 논문에서는 연속음성 인식 시스템 구성을 위한 HMM Word Spotting기법을 검토하였다. 실험에 사용한 HMM Word Spotting기법은 $O(n)$ DP법과 OPDP법이다. 인식 시스템은 파라미터로 펄스스트림만을 사용한 경우와 동적 파라미터인 회귀계수를 결합한 경우의 2종류이며 인식 알고리즘은 $O(n)$ DP법과 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 나눌 수 있다. 또한 인식 단위는 음절과 단어가 혼합된 형태이고 학습은 모두 음절단위로 실시하였으며 연속음성 25문장에 대하여 $O(n)$ DP법과 OPDP법의 인식결과를 비교하여 연속음성 인식에 구문제어 효과를 검증하였다.

실험 결과 평균 인식률이 $O(n)$ DP의 경우 각각 90.6%, 90.9%, OPDP의 경우 각각 98.4%, 98.6%로 유한 상태 오토마타에 의한 구문제어를 이용한 경우 평균 7.5%의 인식률이 향상되었다.

I. 서론

최근 음성인식 연구는 음향학적 인식 단계에서 언어·지식과의 관계가 강조되어 음향 신호처리와 언어·지식 레벨에서의 정보처리가 통합되는 경향에 따라 음성인식에 있어 추도적 역할을 해온 DP매칭 방법의 단점들을 보완할 수 있는 방법으로 HMM법이 각광 받기 시작했다. 이와 같은 배경에서 연구된 HMM법은 음성의 변동을 통계적으로 처리하고 이 통계량을 확률 형태의 모델에 반영하여 음성을 인식하는 방법이며 HMM을 음성 인식에 적용할 경우 확률 모델을 사용하기 때문에 개인차나 조음 결함의 영향 등에 의한 음성 패턴의 변동을 반영하기 쉽고 확률 통계론에 의한 이론적 근거가 쉬우며 음소나 음절 단위의 모델을 단어, 문장 등의 단위로 쉽게 확장할 수 있다는 장점이 있으나, 모델 구조의 결정에는 시행착오나 경험에 의존하는 요소가 강하며, 학습에는 다량의 샘플 데이터와 계산 능력이 필요하고, 음성의 과도적 정보를 정사하는 경향이 있기 때문에 패턴의 시간적 상간의 표현력이 부족하다는 결점도 존재한다. [8][9]

현재의 HMM의 연구는 이러한 결점에 대한 대책으로 음성의 동적 특성을 표현할 수 있는 회귀계수와 지속시간 확률 등을 파라미터로 추가하고, 음향학적 신호 처리와 언어·지식 단계의 정보처리가 통합되는 경향에 따라 신경회로망, 청각·지각 모델 및 언어 처리의 기술을 HMM에 적용하는 연구가 활발히 진행되고 있다.

본 논문에서는 이러한 연구 결과들을 고려하여 연속음성 HMM 모델에 음성의 동적 특성을 표현할 수 있는 회귀계수를 파라미터로 추가한 연속음성분포형 HMM 모델을 구성하고 이 모델을 $O(n)$ DP법과 또한 구문상의 제어를 실시하여 소규모 연속음성 인식 작업을 기술하는데 알맞은 언어 모델인 유한 상태 오토마타(Finite State Automata)를 이용하여 HMM 연속음성 인식시스템에 적용하여 인식성능을 비교하였다.

II. HMM의 모델링

2.1 연속 분포 HMM

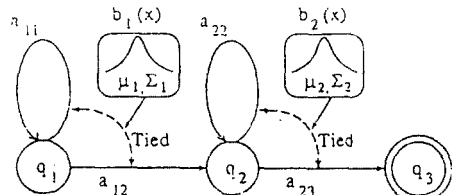


그림 1. 연속분포 HMM의 예

Left-to-Right형 HMM은 그림 1과 같은 확률 유한 오토마타로 정의된다. HMM을 이용해서 음성인식을 실시할 때에는 미리 필요한 표준편차 수만큼 모델을 학습해 놓고 미지의 입력패턴에 대해서 그 출력확률이 최대인 표준편차를 인식결과로 하는 것이다. 이산 분포 HMM의 경우 상태 i 에서 j 로의 전이확률 a_{ij} 및 그 전이경로에서 심벌 k 의 출력확률 b_{ik} 를 학습 데이터로부터 구하기 위한 알고리즘을 Baum-Welch 알고리즘이라 하고 그 내용은 다음과 같다. [4][7][8]

상태수를 N , T 를 심벌 계열이라 하고 관측계열 o_1, o_2, \dots, o_T 를 생성해서 상태 i 에 이르는 확률을 전향확률 $\alpha(i, t)$ ($i=1, 2, \dots, N; t=1, 2, \dots, T$)로 정의하고, 상태 j 에서 시작되는 상태연이에 의해서 $o_{t+1}, o_{t+2}, \dots, o_T$ 가 생성되는 확률을 후향확률 $\beta(j, t)$ ($j=1, 2, \dots, N; t=T, T-1, \dots, 0$) 및 모델 M 이 심벌 계열 $o = o_1 o_2 \dots o_T$ 를 출력하는 확률을 $P(o | M)$, 상태 i 에서 상태 j 로의 전이가 시각 t 에서 발생 할 확률을

$$\gamma_r(i, j) = \frac{a(i, t-1) a_{ij} b_{ij}(o_t) \beta(j, t)}{P(o_t | M)} \quad (1)$$

로 정의하면 전이확률 a_{ij} 와 출력확률 재추정치는 각각

$$a_{ij} = \frac{\sum_t \gamma_r(i, j)}{\sum_t \sum_j \gamma_r(i, j)} \quad (2)$$

$$b_{ijk} = \frac{\sum_{t: o_t=k} \gamma_r(i, j)}{\sum_t \gamma_r(i, j)} \quad (3)$$

로 된다. 이러한 파라미터 재추정을 반복하면 부분적으로 최적치에 수렴하는 것이 증명되어 있다.

식(3)과 같이 HMM의 출력확률분포가 이산분포로써 취급된 경우 음성특징 파라미터는 벡터 양자화에 의해 코드북 벡터 번호로 부호화 되어 심벌 계열로 된다. 그러나 벡터 양자화에 의한 양자화 오차를 피할 수 없고 벡터요소간의 상관의 정보가 손실되기 때문에 출력확률분포를 연속분포로 하면 인식이 향상되는 것으로 보고되어 있다.

출력벡터 o_t 가 n차원의 정규분포에 따른다고 가정할 수 있는 경우 출력확률밀도 함수는

$$b_{ij}(o_t, \mu_{ij}, \Sigma_{ij}) = \frac{\exp\{-\frac{1}{2}(o_t - \mu_{ij})^T \Sigma_{ij}^{-1} (o_t - \mu_{ij}) / 2\}}{(2\pi)^{n/2} |\Sigma_{ij}|^{1/2}} \quad (4)$$

로 주어진다. 여기서, μ_{ij} 는 출력 벡터의 평균치, Σ_{ij} 는 공분산행렬, t 는 전치, -1 은 역행렬을 나타낸다.

이때

$$\gamma_r(i, j) = \frac{a(i, t-1) a_{ij} b_{ij}(o_t, \mu_{ij}, \Sigma_{ij}) \beta(j, t)}{P(o_t | M)} \quad (5)$$

로 정의하면, 전이확률 추정식은 식(2)와 같고, μ_{ij} , Σ_{ij} 의 추정식은 다음 식으로 주어진다.

$$\mu_{ij} = \frac{\sum_t \gamma_r(i, j) o_t}{\sum_t \gamma_r(i, j)} \quad (6)$$

$$\Sigma_{ij} = \frac{\sum_t \gamma_r(i, j) (o_t - \mu_{ij})(o_t - \mu_{ij})^T}{\sum_t \gamma_r(i, j)} \quad (7)$$

2.2 동적 특징 파라미터를 이용한 HMM

HMM에서는 음성의 어느 짙은 정상구간에 HMM의 1개의 상태가 대응한다고 하는 성질이 있다. 음성 패턴의 동적 특징을 무시하지 않기 위해서는 상태 수를 프레임수에 가깝게 증가시켜 자기 전이 루프의 전이 확률을 감소시켜야 하나 이는 계산상 부리가 따른다. 따라서 음성인식의 동적 특징 파라미터로써 회귀계수에 의한 스펙트럼의 동적특징량의 이용을 고려한다. 음성의 정적특징량(스펙트럼계수) 벡터의 각각의 차원에 대해서 식(9)로 계산되는 시간축방향의 선형 회귀계수를 파라미터로 결합한다. 학습 인식에 있어서 2개의 동적 특징 파라미터인 정적 특징량과 회귀계수의 각각에 대해서 식(5)로 출력확률을 계산해 놓고 식(8)과 같이 이들의 곱을 벡터에 대한 출력확률로 계산한다. 여기서 정적특징량과 회귀계수의 상관은 고려하지 않았다.[3][8][9]

$$b_{ij}(o_t) = b_{ij}^p(o_t) \cdot b_{ij}^r(o_t) \quad (8)$$

시계열 $x(t)$ 에 있어서 시간 t 를 중심으로한 $2n+1$ 폭의 선형 회귀계수 $\Delta x(t)$ 는 다음 식으로 계산된다.

$$\Delta x(t) = \frac{\sum_{i=-n}^n i \cdot x(t+1)}{\sum_{i=-n}^n i^2} \quad (9)$$

t : 프레임, n : 회귀계수 폭, i : 회귀계수 차수

III. HMM에 의한 연속음성인식

3.1 O(n)DP법에 의한 연속음성인식

O(n)DP법은 입력된 음성 데이터의 시계열에 대해 우도가 최대가 되도록 단어(음절) HMM을 연결하고 연결된 단어(음절) HMM의 모델 심벌을 출력하는 방법이다. 이때 우도의 계산과 최적인 단어 계열을 구할 때에는 Viterbi 알고리즘을 이용한 다. 그림 2에서 모델 N, 상태가 j라 할 때 입력음성 데이터의 시계열 t를 1에서 T까지 증가시키면서 각 프레임에서의 모델 n, 상태 j에서의 최대누적 확률을 $f^n(j, t)$ 에 기억시키고 $f^n(j, t)$ 에 대한 백포인터(Back Pointer) $B^n(j, t)$ 를 구한다. 모든 모델 n에 대해서 $f^n(j, t)$ 가 구해지면 누적확률이 최대인 $f^n(j, t)$ 를 결정하고 그 확률을 t프레임까지 단어(음절)열의 최대 누적확률로 L(t)에 기억시킨다. 이와 같은 과정을 모든 격자점에 대해 T 프레임까지 반복하고 역추적(Back tracking)을 통하여 인식 결과를 출력하게 된다.

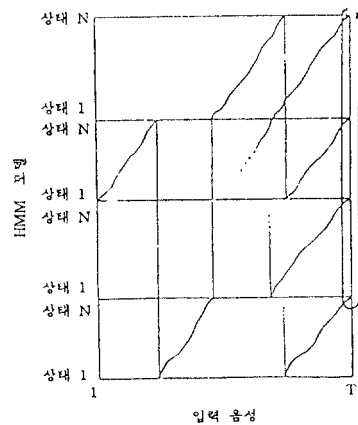


그림 2. O(n)DP법의 최적 패스 검색

3.2 구문제어 One Pass DP법에 의한 연속음성 인식

One Pass DP법은 유한상태 오토마타(Finite State Automata)에 의한 구문제어를 통하여 효율적인 탐색을 실시하는 연속음성 인식 알고리즘이다. 구문제어 One Pass DP법에 의한

연속음성 인식 알고리즘은 $O(n)$ DP법과 기본적으로 같으나 오토마타 제어루트가 추가되어 HMM 모델의 모든 단어가 최적 단어열 검색에 이용되지 않고 현재 진행중인 오토마타 상태 번호까지 구성할 수 있는 부분 문에 등록된 단어만을 검색하는 방법이다. 그림 3은 실험에 사용된 로보트 제어 명령문을 생성하는 유한 상태 오토마타를 나타냈다.

입력 음성 시계열 데이터 각 프레임에서 유한 상태 오토마타의 상태 q 에 도달할 수 있는 단어열은 HMM 모델을 연결하여

$$M = M_{w_1}(1) \otimes M_{w_2}(2) \otimes \dots \otimes M_{w_n}(x)$$

과 같이 구할 수 있다. 즉 그림 3의 예에서 오토마타 상태 ③에 도달할 수 있는 부분 문은 "앞으로", "뒤로" 2가지 부분 문을 구성할 수 있고 각각에 대해서 입력 패턴이 출력하는 확률이 최대가 되는 단어열 $w_q(1), w_q(2), \dots, w_q(x)$ 를 구한다. 단, HMM의 연결 \otimes 은 앞 HMM의 최종 상태와 다음의 HMM의 초기상태를 null(Null)천이로 연결한 것이다. 이와 같은 과정을 유한 상태 오토마타의 모든 상태 q 에 대해서 실시하고 최종 상태에 이르러 모든 단어열 중 누락 대수 확률이 가장 큰 단어열을 백 트래킹하여 인식결과를 출력한다.

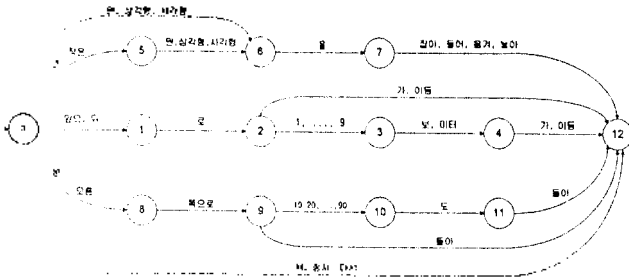


그림 3. 표 2.의 로보트 제어 명령문에 대한 FSN

그림 4는 본 연구의 OPDP법에 의한 연속음성 인식 시스템의 구성도이다. 시스템은 음성 데이터베이스 작성부, HMM모델 학습부, OPDP법에 의한 인식부로 대별할 수 있다.

음성 데이터베이스 작성부는 녹음된 음성을 샘플링, 특징과 파라미터 추출, 라벨링(Labeling)에 의해 음성 데이터베이스를 작성하고 학습과 인식 처리부에 음성 데이터를 제공한다. HMM 모델의 학습은 Baum-Welch 알고리즘에 의해 음절단위로 학습되며 단어사전과 연결기에 의해 단어 또는 음절 단위로 학습된 모델이 저장된다. 음성인식 시스템에서 이러한 언어 모델을 사용하는 목적은 언어적인 제약을 통해서 최적인 단어 열을 탐색하는 것이다.

IV. 인식 실험 및 결과 고찰

4.1. 음성자료 및 분석방법

본 실험에 이용된 음성 데이터의 분석조건은 표 1과 같다.

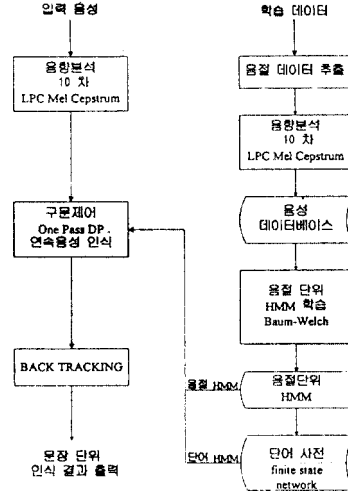


그림 4. 연속음성 인식 시스템의 구성

표 1. 음성 데이터의 분석조건

A/D 데이터	10 kHz, 12 Bit
고역강조	1 차차분
프레임 간격	5 ms
분석창 길이	10 ms
특징 파라미터	LPC Cepstrum(14차) → LPC Melcepstrum(10차)
회귀계수	10 차, 45 ms 폭

학습 및 인식에 사용된 연속음성은 로보트 제어 명령에 관한 문장을 그림 3과 같은 FSN(Finite State Network)로 구성하고 이 오토마타에서 구성할 수 있는 문장 가운데 25 문장(표 2)을 선택하여 5명의 남성 화자가 5회씩 발생한 연속음성이다. 학습에 사용된 음절은 연속음성 중에서 시찰에 의해서 추출하였으며 각 화자의 5회 발생 가운데 2회분을 이용하였으며 각 음절의 발생 빈도가 일정하지 않기 때문에 하나의 음절당 학습 샘플의 수도 일정하지 않다. 인식실험에서는 각 화자가 발생한 5회의 연속음성 모두를 시험 패턴으로 사용하였다.

표 2. 실험에 사용된 로보트 제어 명령문

1) 정지	14) 삼각형을 들어
2) 서	15) 사각형을 옮겨
3) 다시	16) 원을 잡아
4) 앞으로 가	17) 삼각형을 옮겨
5) 뒤로 이동	18) 큰 삼각형을 잡아
6) 앞으로 5미터 이동	19) 큰 원을 놓아
7) 앞으로 4보 가	20) 큰 사각형을 들어
8) 뒤로 2보 가	21) 큰 사각형을 옮겨
9) 뒤로 3미터 가	22) 작은 사각형을 들어
10) 왼 쪽으로 10도 돌아	23) 작은 삼각형을 잡아
11) 오른 쪽으로 45도 돌아	24) 작은 원을 옮겨
12) 왼 쪽으로 돌아	25) 작은 원을 놓아
13) 오른 쪽으로 돌아	

4.2. 모델 구조 및 학습

본 실험에 이용한 음절 HMM의 구조를 그림 5에 나타냈다. 각 음절 HMM 모델은 모두 5상태 4출력 분포 HMM이며 출력확률은 연속 출력확률분포 HMM을 이용했다.

본 연구의 연속음성 인식 실험은 인식 때에는 음절이나 단어 단위로 실시하고 있으나 HMM 모델의 학습은 모두 음절 단위로 실시한다. 단어단위 인식의 경우 이미 학습된 음절단위 HMM 모델이 앞 음절의 최종상태와 후속 음절의 초기 상태가 널(null) 현이로 연결되도록 하는 연결기와 단어사전을 작성하여 자동 구성되도록 하였다.

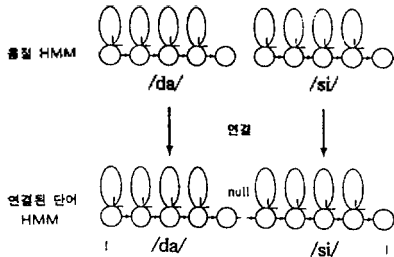


그림 5. 음절 HMM과 단어 HMM의 구조

학습의 과정을 그림 6에 나타낸다. 학습전의 초기 파라미터는 모든 표준화된 학습용 데이터를 시간축에 따라서 출력확률분포수로 균등하게 나누고 각 구간의 특징량 벡터로부터 구한다. 그후 새로운 학습 샘플이 가해지면 이미 구해진 μ (평균), Σ (공분산)과 관측 벡터 o_t 사이에서 Baum-Welch 학습 알고리즘에 의해서 새로운 출력 확률을 구하고 이 출력 확률에 의해 다시 새로운 μ (평균), Σ (공분산)이 구해진다. 본 실험의 학습에서는 이러한 과정을 1샘플당 10회 반복 학습되고 다시 새로운 샘플이 가해져서 학습이 계속된다. 학습에는 Baum-Welch의 재추정 알고리즘을 사용해서 초기 HMM을 10회 반복해서 학습한다.

4.3. 인식결과 분석 및 고찰

본 논문의 실험에서는 멜렙스트림 만을 파라미터로 사용한

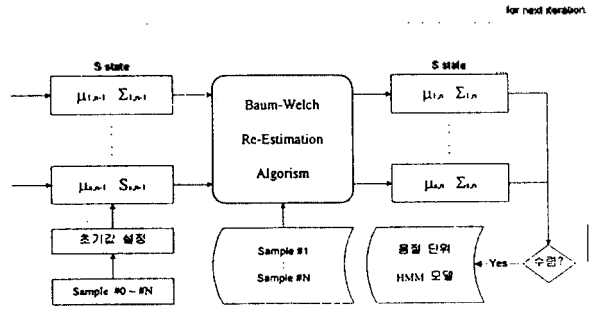


그림 6. Baum-Welch 알고리즘에 의한 반복학습의 과정

연속분포 HMM을 기본 시스템으로 구성하고 등적파라미터인 회귀계수를 결합한 경우에 대하여 25문장의 로버트 제어명령문을 $O(n)$ DP법과 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 인식하여 연속음성 인식에서 구문제어의 효과를 검증하였다.

표 3은 표 2에 나타난 25개의 문장을 그림 3과 같은 FSN로 구성된 구문제어에 의한 One Pass DP법으로 인식한 결과이다. 인식결과 $O(n)$ DP의 경우 90.9%, One Pass DP의 경우 98.6%의 인식률을 나타냈다. 여기서 치환에 의한 오인식은 FSN 상태 ③과 ④의 숫자를 인식에서 모두 발생하였으며 그 밖의 상태에서는 비교되는 패턴의 수가 최대 4개 이하로 제한되는 단순한 문장이므로 오인식이 발생하지 않는 것은 당연한 결과이다. 회귀계수를 이용한 경우 삼입에 의한 오인식으로 인식률의 향상이 0.25% 정도도 미미 하였으나 이는 간단한 전 처리나 후처리에 의해서 개선 될 수 있다.

V. 결론

본 논문에서 구정한 연속음성 인식 시스템은 연속분포 HMM과 회귀계수를 파라미터로 이용한 경우의 2종류이며 인식 알고리즘에 따라 $O(n)$ DP법과 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 나눌 수 있다. 또한 인식의 단위는 음절과 단어 단위가 혼합된 형태이며 학습은 모두 음절단위로 실시하였다. 각 음성 인식 시스템의 성능을 비교하기 위하여 연속음성 25문장을 인식 대상으로 하였다.

실험 결과 $O(n)$ DP의 경우 평균 90.9%, 유한 상태 오토마타에 의한 구문제어를 이용한 경우 평균 98.6%로 7.5%의 인식률

표 3. $O(n)$ DP 과 One Pass DP 의 인식률 비교

인식당인 문장의 구분	O(n) DP										One Pass DP													
	MCEP					MCEP + RGC					MCEP					MCEP+RGC								
	음 결 수	오 인 식 입 력	오 인 식 타 락	오 인 식 치 환	인 식 율	음 결 수	오 인 식 입 력	오 인 식 타 락	오 인 식 치 환	인 식 율	음 결 수	오 인 식 입 력	오 인 식 타 락	오 인 식 치 환	인 식 율	음 결 수	오 인 식 입 력	오 인 식 타 락	오 인 식 치 환	인 식 율				
M1	728	55	2	14	71	90.2	728	59	1	18	78	89.3	728	.	.	10	10	98.6	728	.	.	9	9	98.8
M2	711	33	11	20	64	91.0	711	26	9	23	58	91.8	711	.	.	14	14	98.0	711	.	.	10	10	98.5
M3	730	44	11	13	68	90.7	730	39	4	18	61	91.6	730	.	.	11	11	98.5	730	.	.	11	11	98.5
평균	723	44	8	16	68	90.6	723	41	5	20	66	90.9	723	.	.	12	12	98.4	723	.	.	10	10	98.6

• 인식률=(전음결수-삼입-타락-치환)/전음결수*100
• MCEP: 멜렙스트림 계수, RGC:회귀계수

이 향상되어 구문제어의 효과가 대단히 크게 나타남을 알 수 있었으며 항공탑승권이나 철승차권 예약, 질문 응답에 의한 북경 지역의 관광 안내와 같은 소규모 연속음성 시스템에 적용할 수 있을 것으로 기대된다. 그러나 이러한 구문제어에 의한 연속음성 시스템은 대어휘 연속음성 인식에는 FSN의 상태수가 지나치게 증가하여 적용될 수 없으며 언어 처리 모델을 한국어 연속음성에 적용하기 위해서는 한국어 구문구조에 대한 연구가 필요하다.

参 考 文 献

- (1) 김종호, "마크코프 체인", 동국대학교 출판부, 1987
- (2) 조경호 외, "상태의占有時間 情報를 包含하는 Hidden Markov Model", 한국 전자공학학회지, 제 27 권 제 4호 1990
- (3) 은종관 외, "연속분포 HMM을 이용한 한국어 연속 음성인식 시스템 개발", 한국음향학회 VOL. 13 NO. 1 PP.24-31, 1994
- (4) J. K. Baker, "The dragon system-an overview", IEEE Trans. Acoust. Speech & Signal Process., ASSP-23, PP.24-29, 1975.
- (5) L. R. Rabiner, et al, "On the application of vector quantization and hidden Markov models to speaker independent, isolated word recognition", Bell Syst. Tech. J., 62, PP.1075-1105, 1983.
- (6) K-F. Lee and H-W. Hon, "Large-vocabulary speaker-independent continuous speech recognition using HMM", Proc. ICASSP, PP.123-126, 1988.
- (7) 中川聖一, "確率モデルによる音聲認識", 電子情報通信學會編, 1899.
- (8) 中川聖一, "連続出力分布型HMMによる日本音聲認識" 音響學會 論文誌 Vol.46, PP.486-496, 1990.
- (9) 박창호, 김상범, 허강인, "CHMM을 이용한 음소화 단어 음성 인식에 관한 연구", 한국통신학회 부산·경남지부 학술 논문 발표회 논문집, 제 1집, PP. 67-71, 1994.
- (10) 이종진, 안경영, 허강인, "Hidden Markov Model을 이용한 한국어 연속음성 인식", 한국통신학회 부산·경남지부 학술 논문 발표회 논문집, 제 1집, PP. 72-76, 1994.