

이산 지속시간제어 연속분포 HMM을 이용한 연속 음성 인식

Korean Continuous Speech Recognition Using Discrete Duration Control Continuous HMM

이 종 진*, 김 수 훈**, 허 강 인**

(Jong Jin Lee*, Soo Hoon Kim**, Kang In Hur**)

요 약

본 논문에서는 연속분포 HMM에 이산 지속시간제어와 회귀계수를 파라미터로 추가한 이산 지속시간제어 연속분포 HMM 모델을 이용하여 한국어 연속음성 인식 시스템을 구성하였다. 또한 25 문장의 로봇트 제어명령문을 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 인식 실험을 실시하였다.

4연 숫자음에 대한 인식 실험에서 이산 지속시간 제어와 회귀 계수를 포함한 경우 평균 93.8%의 인식율을, 포함하지 않은 경우 80.7%의 인식율을 얻었다. 로봇트 제어 명령문의 인식에서는 구문제어를 실시하지 않은 경우 평균 90.9%, 유한 상태 오토마타에 의한 구문제어를 이용한 경우 평균 98.4%로 7.5%의 인식율이 향상되었다.

ABSTRACT

In this paper, we report the continuous speech recognition system using the continuous HMM with discrete duration control and the regression coefficients. Also, we do recognition experiment using One Pass DP method (for 25 sentences of robot control commands) with finite state automata context control.

In the experiment for 4 connected spoken digits, the recognition rates are 93.8% when the discrete duration control and the regression coefficients are included, and 80.7% when they are not included.

In the experiment for 25 sentences of the robot control commands, the recognition rate are 90.9% when FSN is not included and 98.4% when FSN is included.

I. 서 론

음성에 의한 인간-기계의 인터페이스는 속도가 빠르고 특별한 훈련없이 이루어질 수 있다는 점에서 많은 이점을 가지고 있고 음성인식 기술의 확립은 각각의 중요한 연구과제가 되고 있다. 특히 최근 10 여년

간의 급속한 기술 개발로 미국, 일본 등지에서는 격리 단어를 인식할 수 있는 상업용 제품이 나와 있으며 연속단어를 인식할 수 있는 시스템들도 연구실에서 개발되고 있다. 그러나 현재 제한 없는 자연 언어 음성인식(블록정화자 대어휘 연속음성 인식 및 이해)은 불가능한 상태이고 화자, 어휘수, 발성내용, 환경 등의 제한 하에서 연구가 이루어지고 있다.

기존의 음성 인식 방법인 DP매칭법은 시계열 패턴

*장안전문대학 전자계산과

**동아대학교 전자공학과

접수일자: 1994년 9월 24일

의 시간축상에서의 비선형 신축을 허용하는 패턴 조희 알고리즘이다. 이것은 단어(또는 음운, 음절)에 대해서 표준적인 시계열 패턴을 표준패턴으로 하고 입력된 시계열 패턴을 비선형으로 신축해 가면서 조희하는 방법이다.

이 방법은 시계열 패턴의 시간적 구조의 변동을 잘 흡수할 수가 있지만 화자의 개인차 등에 기인하는 스펙트럼 그 자체 변동에 대해서는 처리가 어렵다.

한편, 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 결합의 영향 등에 의한 음성 패턴의 변동을 반영하기 쉽고 확률 통계론에 의한 이론적 전개가 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 경시하는 경향이 있기 때문에 패턴의 시간적 상관의 표현력이 부족하다는 결점도 존재한다.[8][9]

현재 연속음성 인식을 위한 HMM의 연구는 이러한 결점에 대한 대책으로 음성의 동적 특징을 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.

한편 국내에서도 연속분포 HMM을 이용한 연속음성 시스템이 개발되었고, 이산분포 HMM에 동적인 파라미터인 회귀계수를 결합하거나 지속시간을 확률적으로 제어하면 인식이 크게 향상된다는 연구보고도 있었다.[2][3][4]

본 논문에서는 이러한 연구 결과들을 종합적으로 고려하여 기본적인 연속분포 HMM 모델에 음성의 동적 특징을 표현할 수 있는 회귀계수와 지속시간 제어를 파라미터로 추가한 이산분포 지속시간제어 연속출력분포형 HMM 모델을 구성하고 이 모델을 O(n)DP법을 이용하여 연속 음성인식에 적용하였다. 또한 구문상의 제어를 실시하여 소규모 연속음성 인식 작업을 기술하는데 알맞은 언어 모델인 유한상태 오토마타(Finite State Automata)를 연속출력 HMM 연속음성 인식시스템에 적용하여 인식성능을 비교하

였다.

II. HMM의 모델링

마코프 과정은 시점 t 에서 $X(t)$ 의 값이 알려져 있을 때 t 보다 이후의 시점 $s(>t)$ 에서의 $X(s)$ 의 확률법칙이 t 보다 이전의 시점 $u(<t)$ 에서의 $X(u)$ 의 값에 관계없는 확률과정이다. 이 경우 현재의 상태가 정확히 알려졌을 때 장래의 확률법칙이 과거의 이력에 영향을 받지 않는 것을 나타낸다. 이와 같은 마코프 과정은 조건부 확률에 의해서 다음과 같이 정의된다.[1]

확률과정 $\{X(t)\}$ 가 있을 때 시점 $t_1 < t_2 < \dots < t_n < t$ 에서 관측치들 x_1, x_2, \dots, x_n , 이라 할 때

$$\begin{aligned} P(X(t)=x | X(t_1)=x_1, X(t_2)=x_2, \dots, X(t_n)=x_n) \\ = P(X(t)=x | X(t_n)=x_n) \end{aligned} \quad (1)$$

이 성립하면 이 과정을 단순 마코프과정 또는 1 중 마코프과정이라 한다. 이때 시간 파라미터 t 와 상태공간이 모두 이산적인 것이 마코프 연쇄(Chain)이다. 마코프과정은 상태천이도로 표현할 수 있다. 관측벡터 $x_n, x_{n+1}, \dots, x_{n+m+1}$ 이 가질 수 있는 값을 하나의 상태에 대응시키고 그 상태에서부터 한번 천이 하여 도달하는 상태를 천이 선으로 연결시킨다. 또 이 천이에 대해서 상태천이확률을 표시하고 하나의 천이마다 1 개의 심벌을 출력하도록 나타낸다.

음성인식에서 이용된 HMM은 음성신호는 발생구조의 시간적 변화에 의해서 발생된 신호이므로 프로세스가 한쪽 방향으로만 천이가 가능하도록 제한시킨 left-to-right 모델을 주로 사용하며 초기상태와 최종상태가 설정되어 1회의 상태천이 마다 심벌을 1개씩 출력한다. 다음에 어느 상태로 천이 하는가 또 그 때에 어떤 심벌을 출력하는가는 각각 천이확률과 출력확률에 의해서 통계적으로 결정되어진다.

HMM이 일반적인 마코프 모델과 다른 점은 출력 심벌 계열이 주어져도 상태천이 경로가 한가지로 결정되지 않는다는 점이고 이러한 의미로써 非決定性有限狀態 오토마타(Automata)로 정의할 수 있다.

1. 연속 분포 HMM

left-to-right 형 HMM은 그림 1.과 같은 확률 유

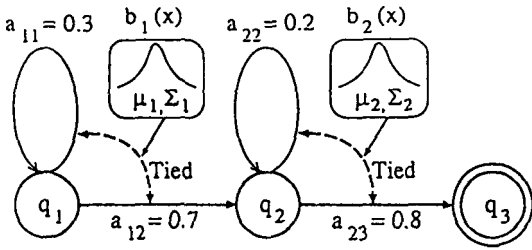


그림 1. 연속분포 HMM의 예
Fig 1. Example of continuous HMM

한 오토마타로 정의된다. HMM을 이용해서 음성인식을 실시하는 실시할 때에는 미리 필요한 표준패턴 수만큼 모델을 학습해 놓고 미지의 입력패턴에 대해서 그 출력확률이 최대인 표준패턴을 인식결과로 하는 것이다. 이산 분포 HMM의 경우 상태 i 에서 j 로의 천이확률 a_{ij} 및 그 천이경로에서 심벌 k 의 출력확률 b_{ijk} 를 학습 데이터에서 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 상태천이에 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 천이가 시작 t 에서 발생할 확률을

$$\gamma_i(i, j) = \frac{\alpha(i, t-1) a_{ij} b_{ij}(o_t) \beta(j, t)}{P(o|M)} \quad (2)$$

로 정의하면 천이확률 a_{ij} 와 출력확률 재추정치는 각각

$$a_{ij} = \frac{\sum_t \gamma_i(i, j)}{\sum_t \sum_j \gamma_i(i, j)} \quad (3)$$

$$b_{ijk} = \frac{\sum_{t: o_t=k} \gamma_i(i, j)}{\sum_t \gamma_i(i, j)} \quad (4)$$

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

식(4)와 같이 HMM의 출력확률분포가 이산분포로써 취급된 경우 경우 음성특징 파라미터는 벡터 양자화에 의해 코드북 벡터 번호로 부호화 되어 심벌 계열로 된다. 그러나 벡터 양자화에 의한 양자화 오

차를 피할 수 없고 벡터요소간의 상관의 정보가 손실되기 때문에 출력확률분포를 연속분포로 하면 인식이 향상되는 것으로 보고되어 있다.

출력벡터 o_t 가 n 차원의 정규분포에 따른다고 가정할 수 있는 경우 출력확률밀도 함수는

$$b_{ij}(o_t, \mu_{ij}, \Sigma_{ij}) = \frac{\exp\{-\frac{1}{2}(o_t - \mu_{ij})' \Sigma_{ij}^{-1} (o_t - \mu_{ij})\}}{(2\pi)^{n/2} |\Sigma_{ij}|^{1/2}} \quad (5)$$

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

$$\gamma_i(i, j) = \frac{\alpha(i, t-1) a_{ij} b_{ij}(o_t, \mu_{ij}, \Sigma_{ij}) \beta(j, t)}{P(o|M)} \quad (6)$$

로 정의하면, 천이확률의 추정식은 식(3)과 같고, μ_{ij}, Σ_{ij} 의 추정식은 다음 식으로 주어진다.

$$\mu_{ij} = \frac{\sum_t \gamma_i(i, j) o_t}{\sum_t \gamma_i(i, j)} \quad (7)$$

$$\Sigma_{ij} = \frac{\sum_t \gamma_i(i, j) (o_t - \mu_{ij})(o_t - \mu_{ij})'}{\sum_t \gamma_i(i, j)} \quad (8)$$

2. 이산확률분포에 의한 지속시간제어 HMM

HMM에서 상태천이확률은 어느 상태가 계속 지속될 것인가 다음 상태로 천이 할 것인가를 나타내는 확률이다. 임의의 상태 i 가 t 시간 지속될 확률은

$$d(t = \text{duration}) = a_{ii}^{t-1} \cdot (a_{ii} - 1) \quad (9)$$

이 되고 시간 t 의 증가와 함께 지수 함수적으로 감소한다. HMM에서 상태 계속시간은 음성 세그먼트의 길이를 나타내며 일반적으로 음성 세그먼트의 길이는 감마 분포나 포아송 분포에 가까운 것으로 알려져 있다. 따라서 기본적인 연속분포 HMM으로는 음성 파형의 과도구간이나 정상구간의 시간 구조를 충분히 표현할 수 없다. 실제 각 상태에서의 지속시간은 음성세그먼트의 발생시간을 나타내는 중요한 정보이므로 인식 시에 이를 고려하는 것이 유리하다. 상태 지속시간제어를 통계적으로 실시하기 위해 a_{ii} 의 자기 루프 천이를 제거하고 대신에 상태 i 가 t 시간 지속될 지속시간의 확률 $d_i(t)$ 을 구하고 이것을 새로운 파라미터로 추가한다. 이때 음성 지속시간확률밀도 분

포를 구하는 방법에 따라 이산분포 지속시간제어 모델과 감마분포나 포아송분포를 이용하는 연속분포 지속시간제어 모델을 구성할 수 있다. 제어의 효과가 가장 우수한 것으로 보고된 감마 분포는 훈련샘플수가 적을 때는 이산 출력 분포보다 정밀한 추정을 할 수 있으나 계산량이 많아진다. 본 연구에서는 훈련 샘플 수가 비교적 충분하다고 가정하고 계산시간이 빠른 장점을 이용하기 위해 이산 분포 지속시간제어 모델만을 고려한다. 단 지속시간 확률은 $\sum_{\tau} d_i(\tau) = 1$ 의 조건을 만족해야 하고 이 파라미터를 도입하면 연속출력 확률밀도분포 HMM의 경우 Baum-Welch의 재추정 알고리즘은 다음과 같이 변화된다.[2][8][9]

$$\alpha(i, j) = \sum_{\tau} \sum_{t \geq \tau} \alpha(j, t - \tau) a_{ij} d_j(\tau) \prod_{n=1}^{\tau} b_{jn}(o_{t+1-n}) \quad (10)$$

$$\beta(i, j) = \sum_{\tau} \sum_{t \geq \tau} a_{ij} d_j(\tau) \prod_{n=1}^{\tau} b_{jn}(o_{t+n}) \beta(j, t + \tau) \quad (11)$$

여기서

$$\gamma(i, j, \tau) = \frac{\alpha(i, t - \tau) a_{ij} d_j(\tau) \prod_{n=1}^{\tau} b_{jn}(o_{t+1-n}) \beta(j, t)}{P(o|M)} \quad (12)$$

로 하면 천이확률 a_{ij} 와 정규분포의 파라미터 μ_{ij}, \sum_{ij} 의 추정식은 각각 다음 식으로 주어진다.

$$a_{ij} = \frac{\sum_{\tau} \sum_{t \geq \tau} \gamma(i, j, \tau)}{\sum_{\tau} \sum_{t \geq \tau} \sum_{i'} \gamma(i', j, \tau)} \quad (13)$$

$$\mu_{ij} = \frac{\sum_{\tau} \sum_{t \geq \tau} \gamma(i, j, \tau) \sum_{n=1}^{\tau} o_{t+1-n}}{\sum_{\tau} \sum_{t \geq \tau} \gamma(i, j, \tau)} \quad (14)$$

$$\sum_{ij} = \frac{\sum_{\tau} \sum_{t \geq \tau} \gamma(i, j, \tau) \sum_{n=1}^{\tau} (o_{t+1-n} - \mu_{ij})(o_{t-n} - \mu_{ij})^t}{\sum_{\tau} \sum_{t \geq \tau} \gamma(i, j, \tau)} \quad (15)$$

또 지속시간확률 $d_j(\tau)$ 의 추정치는

$$d_j(\tau) = \frac{\sum_{\tau} \sum_{t} \gamma(i, j, \tau)}{\sum_{\tau} \sum_{t} \sum_{i'} \gamma(i', j, \tau)} \quad (16)$$

가된다. 식(16)에 의한 추정만으로 학습회수가 진행됨에 따라서 지속시간 확률분포의 차가 크게 나타나 지속시간 확률이 0으로 될 가능성이 있다 따라서 식(17)과 같은 가중치 평균을 이용해서 확률분포의 스프레드를 실시한다.[5][9]

$$d_j(\tau) = \begin{cases} \{2d_j(\tau) + d_j(\tau+1)\}/3 & \text{if } \tau=1 \\ \{d_j(\tau-1) + 2d_j(\tau)\}/3 & \text{if } \tau=\Delta T \\ \{d_j(\tau-1) + 2d_j(\tau) + d_j(\tau+1)\}/4 & \text{else} \end{cases} \quad (17)$$

3. 동적 특징 파라미터를 이용한 HMM

HMM에서는 음성의 어느 짧은 정상구간에 HMM의 1개의 상태가 대응한다고 하는 성질이 있다. 음성 패턴의 동적 특징을 무시하지 않기 위해서는 상태 수를 프레임수에 가깝게 증가시켜 자기 천이 루프의 천이 회수를 감소시켜야 하나 이는 계산상 무리가 따른다.

따라서 음성인식의 특징 파라미터로써 회귀계수에 의한 스펙트럼의 동적특징량의 이용을 고려한다. 음성의 정적특징량(채스트럼계수) 벡터의 각각의 차원에 대해서 식(19)로 계산되는 시간축방향의 선형 회귀계수를 파라미터로 결합한다. 학습 인식에 있어서 2개의 특징 파라미터인 정적특징량과 회귀계수의 각각에 대해서 식(5)로 출력확률을 계산해 놓고 식(18)과 같이 이들의 곱을 벡터에 대한 출력확률로 계산한다. 여기서 정적특징량과 회귀계수의 상관은 고려하지 않았다.[3][8][9]

$$b_{ij}(o_t) = b_{ij}^{cep}(o_t) \cdot b_{ij}^{rc}(o_t) \quad (18)$$

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

$$\Delta x(t) = \frac{\sum_{i=-n}^n i \cdot x(t+1)}{\sum_{i=-n}^n i^2} \quad (19)$$

t : 프레임, n : 회귀계수 폭, i : 회귀계수 차수

III. HMM에 의한 연속음성인식

연속 음성의 경우 단어 경계가 명확하지 않고 단어

내부 뿐아니라 단어 경계 부근의 음이 전후 단어의 영향으로 변화하는 조음결합 현상이 발생하며, 발생 시간이 짧아지고 발음도 불명확해지는 등의 문제점이 추가되어 인식작업은 매우 복잡한 구성을 필요로 하게 된다.

연속음성 인식 기술은 기본적으로 패턴인식의 문제이며 인식하고자 하는 단어(음절)수 만큼 표준 시계열 패턴을 준비하고 이 표준패턴과 입력된 패턴을 비선형으로 조화하는 DP매칭법이 오래 전부터 연구되어 왔다. 현재 확실적인 모델로 표현하는 HMM이 음성인식 기술의 주류를 이루고 있으나 이 HMM 방법도 음향적 우도의 계산에 Viterbi 알고리즘을 이용하고 있다는 점을 제외하면 기본적으로 DP매칭법과 같은 방법이다.[8]

그림 2는 본 연구의 연속음성 인식 시스템의 구성도이다. 시스템은 음성 데이터베이스 작성 부, HMM 모델 학습부, O(n)DP와 One Pass DP 알고리즘에

의한 인식부로 대별할 수 있다.

음성 데이터베이스 작성부는 녹음된 음성을 샘플링, 특징파라미터 추출, 라벨링(Labeling)에 의해 음성 데이터베이스를 작성하고 학습과 인식 처리부에 음성 데이터를 제공한다. HMM 모델의 학습은 Baum-Welch 알고리즘에 의해 음절단위로 학습되며 단어 사전과 연결기에 의해 단어 또는 음절 단위로 학습된 모델이 저장된다. 인식 알고리즘으로는 One Pass DP 알고리즘과 O(n)DP 알고리즘을 이용하였다.

One Pass DP법은 일반적으로 1단 DP(One Stage DP)법으로 부르며 유한상태 오토마타(Finite State Automata: 정규문법)에 의한 구문의 제약을 통하여 효율적인 탐색을 실시하는 연속음성 인식 알고리즘으로 단어사전과 FSN을 참고하여 미리 작성한 상태 정보를 이용하여 인식을 실시한다. O(n)DP법은 One Pass DP법의 오토마타를 1 상태로 감소시켜 구문제약이 없이 최적인 단어계열을 구하는 연속인식 알고리즘이다. 음성인식 시스템에서 이러한 언어 모델을 사용하는 목적은 언어적인 제약을 통해서 최적인 단어열을 탐색하는 것이다.

IV. 인식 실험 및 결과 고찰

1. 음성자료 및 분석방법

본 실험에 이용된 모든 음성 데이터의 분석조건은 표 1.과 같다.

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 4연 숫자음 35개와 이미 잘 알려진 로보트 제어 명령에 관한 문장을 그림 3.과 같은 FSN(Finite State Network)로 구성하고 이 오토마타에서 구성할 수 있는 문장 가운데 25 문장(표 2.)을 선택하여 3명의 남성 화자가 5회씩 발성한 연속음성이다. 학습에 사용된 음절은 연속음성 중에서 시칼에 의해서 추출하였으며 각 화자의

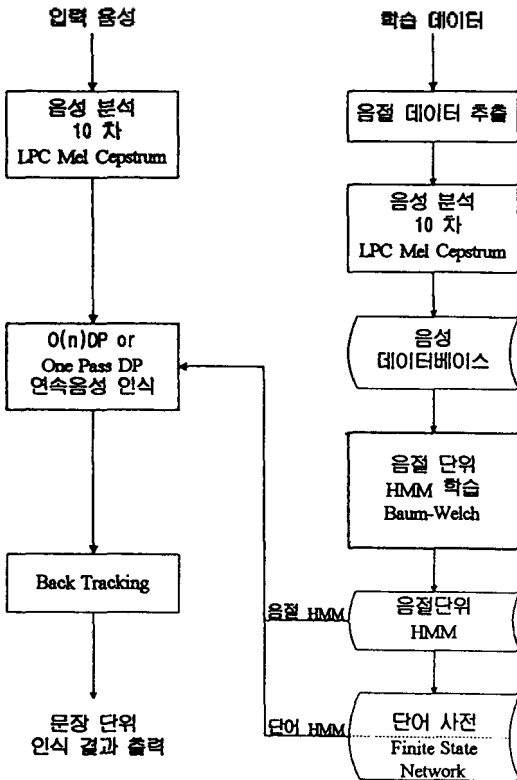


그림 2. 연속음성 인식 시스템의 구성
Fig 2. Block diagram of the continuous speech recognition system

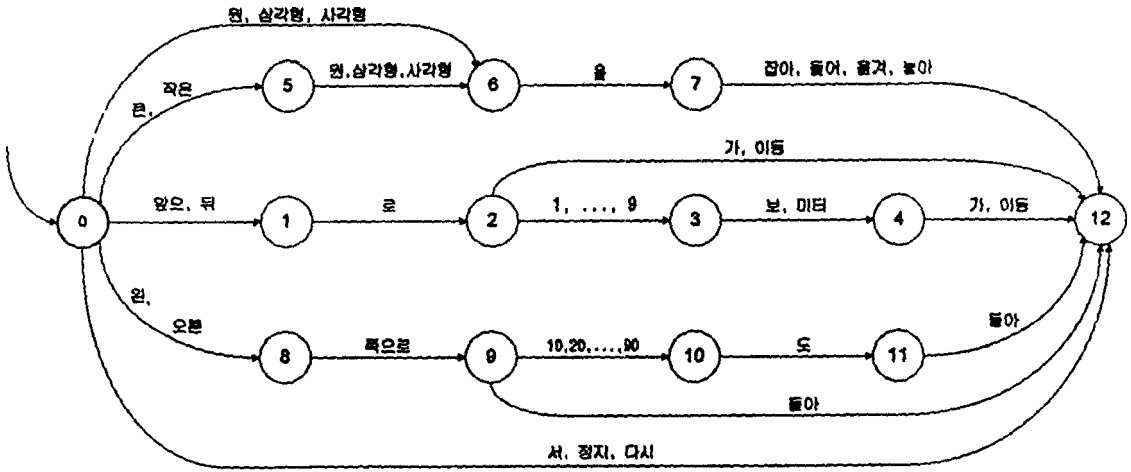


그림 3. 표 2.의 로봇트 제어 명령문에 관한 FSN
Fig 3. Finite State Network of robot control commands

표 2. 실험에 사용된 로봇트 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) 왼쪽으로 10도 돌아	23) 작은 삼각형을 잡아
11) 오른쪽으로 45도 돌아	24) 작은 원을 옮겨
12) 왼쪽으로 돌아	25) 작은 원을 놓아
13) 오른쪽으로 돌아	

5 회 발생 가운데 2 회분을 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.

2. 모델 구조 및 학습

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

본 연구의 연속음성 인식 실험은 인식 때에는 음절 이나 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) 천이로 연결되도록 하는 연결기와 단어사전을 작성하여 자동 구성되도록 하였다.

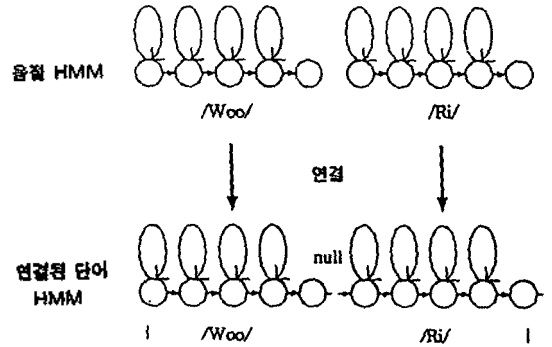


그림 4. 음절 HMM의 구조
Fig 4. Structure of syllable HMM and word HMM

학습의 과정을 그림 5.에 나타낸다. 학습전의 초기 파라미터는 모든 표준패턴의 학습용 데이터들 시간 축에 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 확률

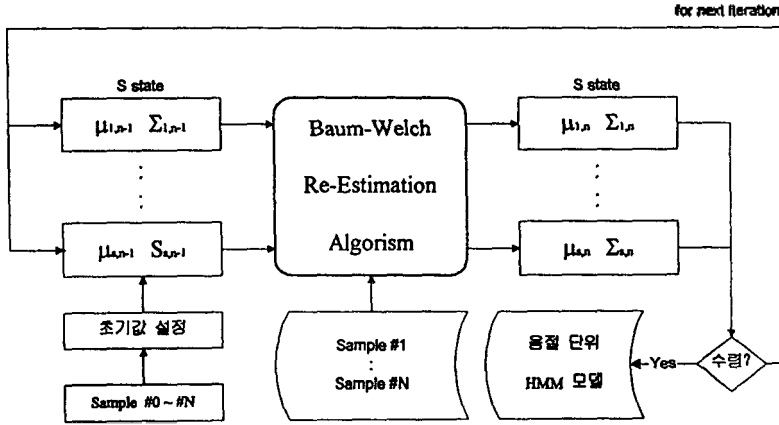


그림 5. Baum-Welch 알고리즘에 의한 반복학습의 과정
 Fig. 5. The training procedure by Baum-Welch algorithm

에 의해 다시 새로운 $\hat{\mu}$ (평균), $\hat{\Sigma}$ (공분산)이 구해진다. 본 실험의 학습에서는 이러한 과정을 1 샘플당 10회 반복 학습되고 다시 새로운 샘플이 가해져서 학습이 계속된다.

이산분포 지속시간 제어 HMM의 경우 천이확률은 모두 1로 설정하기 때문에 그림 4에서 자기 루프 천이는 제거되며 지속시간 확률분포의 초기치는 최대 허용 지속시간의 범위 안에서 균일분포로 준다. 여기서 최대 허용 지속시간은 학습에 이용된 음절 가운데 가장 길게 발생된 음절은 117 프레임 이므로 각 상태당 30 프레임(150 ms)으로 하였다.

학습에는 앞에서 설명한 Baum-Welch의 재추정 알고리즘을 사용해서 초기 HMM을 10회 반복해서 학습한다. 단 지속시간 확률분포는 식(17)에 의해 스무딩된다.

3. 인식결과 분석 및 고찰

본 논문의 실험에서는 멜캡스트럼만을 파라미터로 사용한 연속분포 HMM을 기본 시스템으로 구성하고 동적파라미터인 회귀계수와 지속시간제어를 실시했을 경우의 인식 결과를 4연 숫자음 인식으로 비교실험하였다. 또한 25 문장의 로봇 제어명령문을 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 인식하여 연속음성 인식에서 구문제어의 효과를 검증하였다.

표 3.에는 지속시간 제어와 동적 파라미터인 회귀계수의 사용 유무에 따른 숫자음의 인식 결과를 나타내었다. 인식결과 연속분포 HMM의 경우 평균 82.9%의 인식 결과를 나타내었으며 지속시간제어 연속분포 HMM의 경우 평균 93.8%의 인식율을 나타내었다. 여기서 회귀계수를 기본 연속분포 HMM에 적

표 3. 지속시간제어와 회귀계수에 의한 인식율(숫자음)

인식방법 파라미터 구분	연속분포 HMM												지속시간 제어 연속분포 HMM											
	MCEP						MCEP + RGC						MCEP						MCEP + RGC					
	음 절 수	오 인 식	인 식 율	음 절 수	오 인 식	인 식 율	음 절 수	오 인 식	인 식 율	음 절 수	오 인 식	인 식 율	음 절 수	오 인 식	인 식 율	음 절 수	오 인 식	인 식 율						
M1	700	60	13	58	131	81.1	700	64	9	41	114	83.7	700	8	14	33	55	92.1	700	9	11	28	48	93.1
M2	680	110	11	35	156	77.0	680	111	7	23	141	79.3	680	1	20	24	45	93.4	680	4	14	35	43	93.7
M3	692	64	19	32	115	83.4	692	63	11	24	98	85.8	692	5	19	19	43	93.8	692	5	15	17	37	94.6
평균	691	78	14	42	134	80.7	691	79	9	29	118	82.9	691	5	18	25	48	93.1	691	6	13	27	43	93.8

* 인식율 = (전음절수 - 삽입 - 탈락 - 치환) / 전음절수 * 100
 * MCEP : 멜캡스트럼 계수, RGC : 회귀계수

표 4. O(n) DP법과 구문제어 One Pass DP법의 인식율 비교(로봇제어 명령문)

인식방법 파라미터 구분 화자	O(n) DP										One Pass DP													
	MCEP					MCEP + RGC					MCEP					MCEP + RGC								
	음 절 수	오 인 식				인 식 율	음 절 수	오 인 식				인 식 율	음 절 수	오 인 식				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

* 인식율 = (전음절수 - 삽입 - 탈락 - 치환) / 전음절수 * 100
 * MCEP : 멜렙스트림 계수, RGC : 회귀계수

용한 경우 표 3.과 표 4. 모두 탈락이나 치환에 의한 오인식이 감소하여 전체적으로 인식율이 향상되나 삽입에 의한 오인식 오히려 증가하는 경향이 있음을 알 수 있다. 삽입은 무음 구간이나 단어의 전후에서 주로 발생되었으며 이는 무음구간과 유음 구간 사이에 스펙트럼의 변화가 크게 나타나 회귀계수의 값이 증가하나 스펙트럼의 변화가 미소한 무음 패턴과 매칭이 되지 않기 때문인 것으로 고려된다. 이에 대한 대책으로 회귀계수에 가중치를 가하거나 끝점 검출 알고리즘에 의하여 시퀀스 패턴의 무음구간을 미리 제거한 상태로 인식을 실시하는 방법이 보고되어 있으나 본 실험에서는 실시하지 않았다. 그러나 지속시간 제어를 실시한 경우 각 음절의 지속시간 확률에 의하여 이러한 삽입에 의한 오인식이 대폭 감소하여 10.9% 인식율이 향상되었다.

또한 표 4.는 표 5.에 나타난 25개의 문장을 그림 3과 같은 FSN로 구성된 구문제어에 의한 One Pass DP법으로 인식한 결과이다. 인식결과 O(n)DP의 경우 90.6%, One Pass DP의 경우 98.6%의 인식율을 나타냈다. 여기서 치환에 의한 오인식은 FSN 상태 ③과 ⑩의 숫자음 인식에서 모두 발생하였으며 그 밖의 상태에서는 비교되는 패턴의 수가 최대 4개 이하로 제한되는 단순한 문장이므로 오인식이 발생하지 않는 것은 당연한 결과이다.

를 결합한 경우의 인식 결과를 4연 숫자음 인식으로 비교 실험하였다. 또한 25 문장의 로봇제어명령문을 유한상태 오토마타에 의해 구문제어를 실시한 One Pass DP법으로 인식하여 연속음성 인식에서 구문제어의 효과를 검증하였다.

실험 결과 연속분포 HMM의 경우 평균 82.9%, 지속시간 제어의 연속 분포의 경우 93.8%의 인식 결과로 약 10.9%의 인식 성능이 향상됨을 알 수 있었다. 그러나 회귀계수를 이용한 경우 삽입에 의한 오인식으로 인식율의 향상이 1% 정도로 미미 하였으나 이는 간단한 전 처리나 후처리에 의해서 개선 될 수 있다.

한편, 로봇제어 명령문의 인식에서는 연속분포 HMM의 경우 평균 90.9%, 유한 상태 오토마타에 의한 구문제어를 이용한 경우 평균 98.4%로 7.5%의 인식율이 향상되어 구문제어의 효과가 대단히 크게 나타남을 알 수 있었으며 항공탑승권이나 철승차권 예약, 질문 응답에 의한 특정 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 구문 구조에 대한 연구가 필요하다.

V. 결 론

본 논문에서는 이산분포 지속시간제어 연속분포 HMM에 의한 연속음성 인식 시스템을 구성하고 기본 연속분포 HMM 모델과 동적 파라미터인 회귀계수

참 고 문 헌

1. 김종호, "마르코프-체인", 동국대학교 출판부, 1987
2. 조정호 외, "상태의 占有時間 情報를 包含하는 Hidden Markov Model", 한국전자공학회지, 제27권 제4호,

- 1990
3. 김순협 외, "DHMM을 이용한 한국어 음성인식", 한국음향학회 VOL. 10 NO. 1 PP. 52-60, 1991. NO. 6 PP. 46-52, 1992.
 4. 은종관 외, "연속분포 HMM을 이용한 한국어 연속 음성인식 시스템 개발", 한국음향학회 VOL. 13 NO. 1 PP. 24-31, 1994
 5. J. K. Baker, "The dragon system-an overview", IEEE Trans. Acoust. Speech & Signal Process., ASSP-23, PP. 24-29, 1975.
 6. 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.
 7. K-F. Lee and H-W. Hon, "Large-vocablary speaker-independent continuous speech recognition using HMM", Proc. ICASSP, PP. 123-126, 1988.
 8. 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.
 9. 中川聖一, "連續出力分布型HMMによる日本音韻認識" 音響學會 論文志 Vol. 46, PP. 486-496, 1990.
- ▲이 종 진(Jong Jin Lee) 1958년 8월 20일생
 1982년 2월: 동아대학교 전자공학과 (공학사)
 1986년 2월: 동아대학교 전자공학과 (공학석사)
 1995년 2월: 동아대학교 전자공학과 (공학박사)
 1990년 3월~현재: 장안 전문대학 전산과 조교수
- ▲김 수 훈(Soo Hoon Kim) 1968년 2월 25일생
 1991년 2월: 동아대학교 전자공학과 (공학사)
 1993년 2월: 동아대학교 전자공학과 (공학석사)
 1993년 3월~현재: 동아대학교 전자공학과 박사과정
- ▲허 강 인(Kang In Hur) : 12권 6호 참조