

이산 시간 제어 CHMM을 이용한 한국어 연속 음성 인식에 관한 연구

김상범* 이종진** 허강인*

- * 동아대학교 전자공학과
- ** 장안전문대학 전자계산학과

A Study on Recognition of Korean Continuous Speech using Discrete Duration CHMM.

Sang-Bern Kim Jong-Jin Lee** Kang-In Hur*

- * Dept. of Electronic Eng., Dong-A Univ.
- ** Dept. of Computer Science, Jang Ahn Junior College.

요약

본 연구에서는 확률적 모델을 이용한 HMM(hidden markov model)으로 한국어 연속 음성 인식 시스템을 구성하였다. 학습 모델로서는 양자화 오차가 없는 연속출력 확률밀도를 사용한 연속출력 확률분포 HMM(continuous hidden markov model)과 과도 구간 및 정상 구간의 시간구조를 충분히 표현할 수 없는 것을 계속시간 확률 파라미터를 추가하여 보완한 이산 지속시간 제어 연속출력 확률분포 HMM을 이용하였다. 인식 알고리즘은 시계열 패턴의 시간축상에서의 비선형 신축을 고려한 DP매칭으로서, 음절의 경계를 자동으로 검출하는 O(n)DP를 이용하였다. 실험에서 사용된 연속 음성 데이터는 4연 숫자음과 연속음성 10문장으로 하였다. 인식 실험 결과 4연 숫자음에서 CHMM은 80.7%, DDCHMM은 92.9%의 인식률을 얻었고, 신문 사설에서 발췌한 연속 음성문장의 경우 CHMM 54.2%(중속인 경우 82.1%), DDCHMM에서는 68.9%를 얻어, 시간장 제어를 고려한 DDCHMM이 CHMM보다 높은 인식률을 얻었다.

I. 서론

컴퓨터 및 정보통신 시스템의 급속한 발전과 보급이 활발해짐에 따라 전문가뿐만 아니라 각 분야의 많은 사람들이 이를 이용하게 되었고 인간-기계의 인터페이스 (Man-Machine Interface)에 대한 기대가 높아지게 되었다. 음성에 의한 인간-기계의 인터페이스는 속도가 빠르고 특별한 훈련없이 이루어질 수 있다는 점에서 많은 이점을 가지고 있고 음성인식 기술의 확립은 각각의 중요한 연구 과제가 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(불특정화자 대어휘 연속음성 인식 및 이해)은 불가능한 상태이고 화자, 어휘수, 발성내용, 환경 등의 제한 하에서 연구가 이루어지고 있다.

기존의 음성 인식 방법으로서의 시계열 패턴을 비선형으로 압축하면서 표준 패턴과 유사도를 측정하는 DP매칭법이 주로 연구되어 왔으나 최근 연속음성의 문장인식이나 불특정화자 음성인식에서 인식시간이 길어지는 단점 때문에 음성을

확률적으로 모델링하는 방법인 HMM(Hidden Markov Model)이 연구되고 있다.

본 논문에서는 HMM모델에 연속 출력분포 HMM과 지속 시간 확률 파라미터를 추가한 이산 지속시간 제어 연속출력 확률분포 HMM모델을 구성하였고, 연속음성 인식에는 O(n)DP법을 이용하였다. 단어단위 연속음성 인식의 가장 큰 결점인 표준패턴의 증가에 따르는 학습데이터 양과 학습 시간의 증가를 감소하기 위해 먼저 음절 단위의 HMM을 구성하여 학습하고 이를 단어사전 작성가를 이용하여 NULL 천이로 연결하는 HMM연결법을 이용하여 단어단위의 HMM을 구성하였다.

II. HMM 모델

2.1 연속출력 확률분포 HMM

기존에 사용된 이산출력 확률분포 HMM은 출력 확률분포를 이산분포로써 취급하기 때문에 음성특징 파라미터가 벡터 양자화에 의해 코드북 벡터 번호로 부호화 하여 심벌 계열로 사용되므로서 벡터 양자화에 의한 양자화 오차가 생기게 되는데 이를 없애기 위해 출력확률 분포를 그림 1.에서와 같이 연속분포로써 사용하는 것을 연속 출력분포 HMM이라 한다.

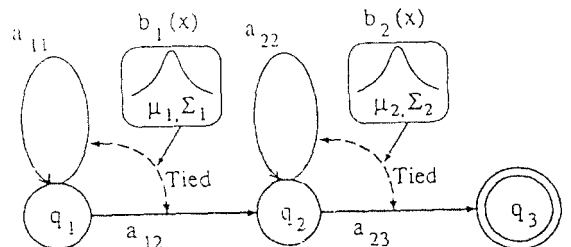


그림 1. 연속출력 확률분포 HMM의 예

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

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

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

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

로 정의하면, 천이확률 a_{ij} 의 추정식, 평균 μ_{ij} 와 분산 Σ_{ij} 의 추정식은 다음 식으로 주어진다.

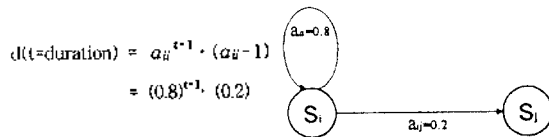
$$a_{ij} = \frac{\sum_t \gamma(i, j)}{\sum_t \sum_j \gamma(i, j)} \quad (3)$$

$$\mu_{ij} = \frac{\sum_t \gamma(i, j) o_t}{\sum_t \gamma(i, j)} \quad (4)$$

$$\Sigma_{ij} = \frac{\sum_t \gamma(i, j) (o_t - \mu_{ij})(o_t - \mu_{ij})^t}{\sum_t \gamma(i, j)} \quad (5)$$

2.2 지속시간제어 IHMM

IHMM에서 상태 천이 확률은 어느 상태가 계속 지속될 것인가 다음 상태로 천이할 것인가를 나타내는 확률이다. 그림 2에서 상태 i 가 t 시간 지속될 확률은



$$d(t=\text{duration}) = a_{ii}^{t-1} \cdot (a_{ij} - 1) = (0.8)^{t-1} \cdot (0.2)$$

그림 2. 천이 확률과 지속시간

이 되고 t 의 증가에 따라 지수 함수적으로 감소한다. IHMM에서 상태 지속시간은 음성 세그먼트의 길이를 나타내며 일반적으로 음성 세그먼트의 길이는 감마 분포나 포아송 분포에 가까운 것으로 알려져 있다. 따라서 기본 IHMM으로는 음성파형의 과도구간이나 정상구간의 시간 구조를 충분히 표현할 수 없다. 실제 각 상태에서의 지속 시간은 음성 세그먼트의 발생시간을 나타내는 중요한 정보이므로 인식 시에 이를 고려하는 것이 유리하다. 상태 지속 시간제어를 통계적으로 실시하기 위해 a_{ii} 의 자기 루프 천이를 제거하고 대신에 상태 i 가 t 시간 지속될 지속 시간의 확률 $d(t)$ 을 구하고 이것을 새로운 파라미터로 추가한 것이 지속시간 제어 IHMM이다.

여기서 지속시간 확률은 $\sum_t d_i(\tau) = 1$ 의 조건을 만족해야 하고 이 파라미터를 도입하면 연속음역 확률밀도분포 HMM의 경우 Baum-Welch의 재추정 알고리즘은 다음과 같이 변화된다.

$$a(i, j) = \frac{\sum_t \sum_{\tau=1}^t a(j, t-\tau) a_{ij} d_i(\tau) \prod_{n=1}^{\tau-1} b_{ii}(o_{t+1-n})}{\sum_t \sum_{\tau=1}^t a_{ij} d_i(\tau) \prod_{n=1}^{\tau-1} b_{ii}(o_{t+1-n})} \quad (6)$$

$$\beta(i, j) = \frac{\sum_t \sum_{\tau=1}^t a_{ij} d_i(\tau) \prod_{n=1}^{\tau-1} b_{ii}(o_{t+1-n}) \beta(j, t+\tau)}{\sum_t \sum_{\tau=1}^t a_{ij} d_i(\tau) \prod_{n=1}^{\tau-1} b_{ii}(o_{t+1-n}) \beta(j, t+\tau)} \quad (7)$$

여기서

$$\gamma(i, j, \tau) = \frac{a(i, t-\tau) a_{ij} d_i(\tau) \prod_{n=1}^{\tau-1} b_{ii}(o_{t+1-n}) \beta(j, t)}{P(o | M)} \quad (8)$$

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

$$a_{ij} = \frac{\sum_t \sum_{\tau=1}^t \gamma(i, j, \tau)}{\sum_t \sum_j \sum_{\tau=1}^t \gamma(i, j, \tau)} \quad (9)$$

$$\mu_{ij} = \frac{\sum_t \sum_{\tau=1}^t \gamma(i, j, \tau) \sum_{o_{t+1-n}} o_{t+1-n}}{\sum_t \sum_{\tau=1}^t \sum_{\tau=1}^t \gamma(i, j, \tau) \tau} \quad (10)$$

$$\Sigma_{ij} = \frac{\sum_t \sum_{\tau=1}^t \gamma(i, j, \tau) \sum_{n=1}^{\tau-1} (o_{t+1-n} - \mu_{ij})(o_{t+1-n} - \mu_{ij})^t}{\sum_t \sum_{\tau=1}^t \sum_{\tau=1}^t \gamma(i, j, \tau) \tau} \quad (11)$$

또 계속시간 확률 $d_i(\tau)$ 의 추정치는

$$d_i(\tau) = \frac{\sum_t \sum_j \gamma(i, j, \tau)}{\sum_t \sum_j \sum_{\tau=1}^t \gamma(i, j, \tau)} \quad (12)$$

로 된다. 식(12)에 의한 추정만으로 학습회수가 진행됨에 따라서 계속시간 확률분포의 차가 너무 크게 나타나기 때문에 다음 식(13)과 같은 가중치 평균을 이용해서 확률분포의 스프레딩을 실시한다.

$$d_j(\tau) = \begin{cases} \{2d_j(\tau) + d_j(\tau+1)\}/3 & \text{if } \tau=1 \\ \{d_j(\tau-1) + 2d_j(\tau)\}/3 & \text{if } \tau = \Delta T \\ \{d_j(\tau-1) + 2d_j(\tau) + d_j(\tau+1)\}/4 & \text{else} \end{cases} \quad (13)$$

III. 연속음성인식 시스템

연속음성은 각각의 단어를 독립적으로 발생할 때에 비해 입력 음성의 패턴이 보다 다양하기 때문에 인식이 어려워진다. 고립 단어 인식에 있어서는 화자에 따른 정도의 구조나 발생 습관의 차이에서 오는 개인차, 전후음운 환경에 의한 조음결합 현상, 발생 시간의 변동 등이 음성 인식을 어렵게 하는 요인이 되고 있으나 연속 음성의 경우 단어 경계가 명확하지 않고 단어 내부 뿐 아니라 단어 경계 부근의 음이 전후 단어의 영향으로 변화하는 조음결합 현상이 발생하며, 단어 단어를 구성하는 음소의 발생시간이 짧아지고 발음도 불명확해지는 등의 문제점이 추가되어 인식작업은 매우 복잡한 구성을 필요로 하게 된다.

연속음성 인식 기술은 기본적으로 패턴 인식의 문제이며 인식하고자 하는 단어(음절) 수만큼 표준 시계열 패턴을 준비하고 이 표준패턴과 입력된 패턴을 비선형으로 조화하는 DP매칭 법이 오래 전부터 연구되어 왔다. 현재에는 확률적인 모델로 표현하는 HMM이 음성인식 기술의 주류를 이루고 있으나 이 HMM 방법도 음향적 우도의 계산에 Viterbi 알고리즘을 이용하고 있다는 점을 제외하면 기본적으로 DP매칭법과 같은 방법이다.

본 연구의 연속음성 인식에 이용된 $O(n)$ DP법은 일반적으로 1단 DP(One Stage DP)법으로 부르며 구분제약이 없이 최적인 단어계열을 구하는 연속인식 알고리즘이다.

3.1 O(n)DP법에 의한 연속음성인식 알고리즘

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 이용한다.

그림 3.에서 모델 N, 상태가 j라 할 때 입력음성 데이터의 시계열 t를 1에서 T까지 증가시키면서 각 t 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 t 프레임까지의 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.

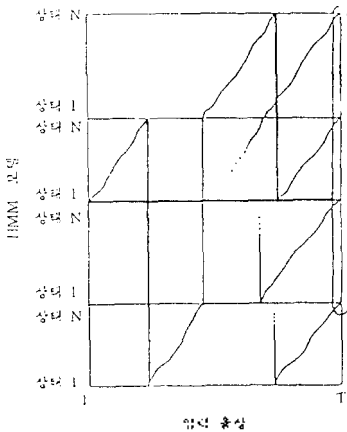


그림 3. O(n)DP법의 최적 피스 검색

3.2 시스템의 구성

그림 4.은 본 연구의 O(n)DP법에 의한 연속음성 인식 시스템의 구성도이다. 시스템은 음성 데이터베이스 작성부, HMM 모델 학습부, O(n)DP법에 의한 인식부로 대별할 수 있다.

음성 데이터베이스 작성부는 녹음된 음성을 12bit, 10kHz로 샘플링하여 펄스 코딩 변조(PCM)와 회귀계수와 같은 특징 파라미터를 추출하고 라벨링(Labeling)과정을 거쳐 음성 데이터베이스를 작성하여 학습과 인식 처리부에 음성 데이터베이스를 제공한다. HMM 모델의 학습은 Baum-Welch 알고리즘에 의해 음절단위로 학습되며 미리 작성된 단어사전과 연결기에 의해 단어 또는 음절단위로 학습된 모델이 저장된다. 연속음성 인식부는 입력된 시험음성과 학습된 HMM 모델을 이용하여 Viterbi 알고리즘을 이용한 O(n)DP법에 의해 최적 단어열을 검색하고 인식결과를 백트래킹에 의해 출력하게 된다.

본 실험에서 O(n)DP법에 의한 연속음성 인식 시스템에 적용된 HMM 모델은 연속분포 HMM 모델과 지속 시간제어 HMM 모델로 구성하였다.

이산 지속시간제어 HMM 모델의 경우 O(n)DP 알고리즘의 출력확률의 계산은 식(6)~식(13)을 이용해서 계산되어진다.

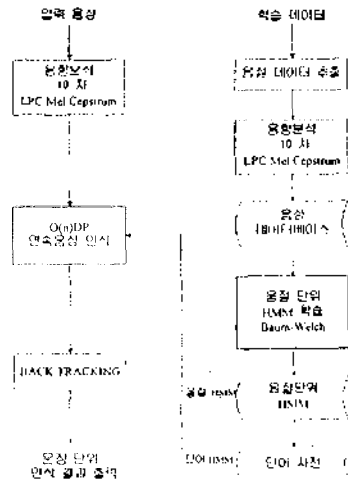


그림 4. O(n)DP법에 의한 연속음성 인식 시스템의 구성도

IV. 인식실험 및 고찰

4.1 음성자료 및 분석 방법

본 실험에 이용된 음성 데이터의 분석은 표 1.과 같이 20대 남성이 발성한 모든 음성을 10kHz로 샘플링하여 분석창길이 10ms, 프레임 간격 5.0ms의 해밍창으로 추출하고 1차 차분에 의해서 고역강조 한 후 14차의 컷스트럼을 구한다. 다시 14차의 컷스트럼 계수는 10차 펄스 컷스트럼 계수로 변환하여 구했다.

표 1. 음성 데이터의 분석조건

A/D 데이터	10 kHz, 12 Bit
고역강조	1 차 차분
프레임 간격	5 ms
분석창	hamming 창
분석창 길이	10 ms
특징 파라미터	LPC Cepstrum(14차) → LPC Melcepstrum(10차)

실험에 사용된 음성 데이터는 표 2, 표 3.에 나타난 바와 같이 35개의 4연 숫자음과 신문 사설에서 발췌된 문장으로 구성된 10문장의 연속 음성으로 구성되어 있다.

여기서 4연 숫자음은 /0/ 에서 /9/ 까지 10종의 각 숫자음 전후에 결합될 수 있는 숫자음의 조합은 100종류가 가능하므로 이를 모두 포함할 수 있도록 선정된 것이며, 35개의 4연 숫자음 가운데에는 각 숫자음이 14회씩 발생되어 총 140개의 숫자음으로 구성되어 있다.

표 3.의 연속 음성은 신문 사설에서 임의로 발췌된 문장이며 따라서 각 음절의 발생 빈도가 일정하지 않고 회화 음성에 가깝게 자연스럽게 발생되었기 때문에 무음 구간이 다수 포함된 음성이다.

3종류의 음성은 모두 3명의 남성 화자에 의해 5회씩 발생되었으며 그 중 2회분의 연속음성에서 목록에 의해서 추출된 음절이 학습에 이용되고 인식시에는 5회분 모두가 시험 피연으로 이용되었다. 따라서 신문 사설의 연속 음성은 각 음절

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	8580	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) 판이란 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 봉애가 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) 전이로 연결되도록 하는 연결기와 단어조건을 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 전이 확률은 모두 1로 설정되기 때문에 그림 6에서 자기 루프 전이는 제거되며 지속시간 확률분포의 초기치는 최대허용 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 자음은 모음의

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

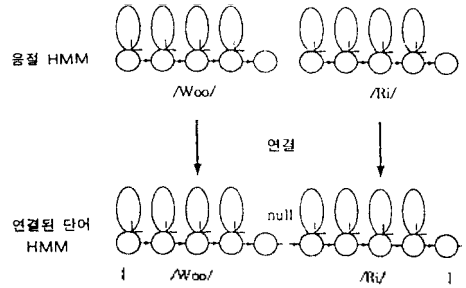


그림 5. 음절 HMM과 단어 HMM의 구조

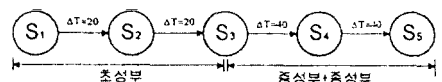


그림 6. 이산 지속시간제어 확률분포 HMM 모델의 구조와 최대 허용시간

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 전이확률)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.

인식 방법		O(n)DP											
HMM모델		CHMM					DDCHMM						
화자	구분	음절수	오인식			인식률 (%)	음절수	오인식			인식률 (%)		
			삽입	탈락	치환			계	삽입	탈락		치환	계
M1		700	60	13	58	131	81.1	700	6	14	33	55	92.1
M2		680	110	11	35	156	77.0	680	1	20	24	45	93.4
M3		692	64	19	32	115	83.4	692	5	19	19	43	93.8
평균		691	78	14	42	134	80.7	691	5	18	25	48	93.1

인식 방법		O(n)DP					
HMM모델		CHMM				DDCHMM	
화자	구분	음절수	오인식			인식률 (%)	
			삽입	탈락	치환		계
종속화자		396	52	2	17	71	82.1
독립화자		563	148	9	102	258	54.2

*인식률=(전음절수-삽입-탈락-치환)/전음절수*100
표 5. CHMM에 의한 화자 종속 및 독립의 인식 결과.

인식 방법		O(n)DP											
HMM모델		CHMM					DDCHMM						
화자	구분	음절수	오인식			인식률 (%)	음절수	오인식			인식률 (%)		
			삽입	탈락	치환			계	삽입	탈락		치환	계
M1		574	148	13	88	259	54.9	574	38	20	88	146	74.6
M2		569	143	11	121	271	52.4	569	53	41	103	197	65.4
M3		546	152	4	88	244	55.3	546	59	14	108	181	66.8
평균		563	148	9	102	258	54.2	563	50	25	100	175	68.9

*인식률=(전음절수-삽입-탈락-치환)/전음절수*100
표 6. CHMM과 DDCHMM에 의한 연속음성의 인식 결과

V. 결론

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

각 연속음성 인식 시스템의 성능을 비교 분석하기 위하여 35종류의 4연 숫자음과 신문 사설에서 발췌한 10문장의 연속음성을 인식 대상으로 하였다.

실험 결과 4연 숫자음의 경우 연속분포 HMM 평균 80.7%, 지속시간 제어 연속 분포 93.1%의 인식 결과를 나타내었으며 약 12.4%의 인식 성능이 향상됨을 알 수 있었다. 그리고 신문 사설의 경우 연속분포 HMM 평균 54.2%, 지속시간 제어 연속 분포 68.9%의 인식 결과를 나타내었으며, 연속분포 HMM에서 화자 종속과 독립의 인식률은 각각 82.1%와 54.2%의 결과를 얻었다.

위의 결과에서 기본적인 연속분포 HMM의 경우 자가 상태에 지속될 확률(자가 루프 천이확률)이 저수 함수적으로 감소하여 각 단어나 음절의 앞뒤에 15 프레임 미만의 짧은 음절이 인식되는 삽입에 의한 오인식이 많은 반면 지속시간 제어의 경우 각 음절의 지속시간 확률에 의하여 이러한 오인식을 대폭 감소시키기 때문인 것으로 분석되어 지속시간 제어의 효과를 증명할 수 있었다. 그러나 지속시간 파라미터가

추가되므로 모델의 학습시간이나 인식시간이 길어지고 계산량이 증가하는 문제가 따르게 된다.

앞으로 실시간 대어휘 연속음성인식에 HMM을 적용하기 위해서는 음소나 보다 작은 단위의 HMM을 구성하여 학습하고 단어단위의 인식을 실시하는 방법이 고려될 수 있으나 이를 위해서는 한국어의 각 음소가 연속 발생될 때 나타나는 조음결합 현상에 대한 기초적인 연구가 필요하다.

參考文獻

- [1] J.K. Baker, "The DRAGON system-An overview," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp. 24-29, Feb. 1975.
- [2] K-F. Lee and H-W. Hon, "Large-vocabulary Speaker-Independent Continuous Speech Recognition using HMM: The SPHINX System", Proc. ICASSP, pp.123-126, 1988.
- [3] 中川聖一, "確率モデルによる音聲認識", 電子情報通信學會誌, 1988.
- [4] 中川聖一, "連続出力分布型HMMによる日本語音聲認識", 音聲學論 論文誌 Vol. 46, pp.486-496, 1990.
- [5] 은종관 외, "연속분포 HMM을 이용한 한국어 연속 음성인식 시스템 개발", 한국음향학회 VOL. 13 NO. 1 pp.24-31, 1994
- [6] 김삼평, 박창호, 허강인, "CHMM을 이용한 음소와 단어 음성 인식에 관한 연구", 한국통신학회 부산경남지부 학술논문 발표회 논문집, 제 1 집, pp. 67-71, 1994.
- [7] 이종진, 안정영, 허강인, "Hidden Markov Model을 이용한 한국어 연속음성 인식", 한국통신학회 부산경남지부 학술논문 발표회 논문집, 제 1 집, pp. 72-76, 1994.