공학 연구논문집 제28권1호 pp. 47 ~ 59울산대학교.
Journal of Engineering Research Vol. 28, No. 1. pp. 47~58, 1997
Unitversity of Ulsan.

HMM을 이용한 음성 자동분절에 관한 연구

김석수 · 안종구
전자공학과

<요 약>

음성 인식기로 사용되고 있는 HMM을 확장된 분류의 음소를 기본 단위로 하여 연속 음성 문장의 음성 분절에 이용하였다. 음소의 분류를 자음, 모음, 묵음, 유음 그리고 비음의 5가지 음소와 정지 자음, 비정지 자음, 모음, 묵음, 유음, 그리고 비음의 6가지 음소로 구성된 두 분류로 나누었다.

HMM의 교육과정에서 사용된 씨앗 문장의 데이터 베이스(DB)는 수작업으로 분절된 30개의 연속 음성 문장이다. 교육된 HMM을 이용하여 분절을 시도하고, 분절된 문장을 데이터 베이스 확장에 사용하였다. 이 과정을 확장된 문장의 수가 120개가 될 때까지 반복하였다. 연속 음성의 분절은 화자 독립인 경우에 대하여 시도하였으며, 두 가지의 분류에 대하여 분절 결과를 비교하였다. 제안된 분류에 대하여는 각 상태의 혼합 계수(mixture coefficient)의 개수 변화에 따라 분절 결과를 비교하였고, 교육된 HMM을 이용하여 자동 분절된 음소의 경계 구간과 수작업으로 분절한 음소의 경계 구간을 비교, 분석하였다.

A Study on Automatic Speech Segmentation Using HMM(Hidden Markov Model)

Seok Soo Kim · Chong Koo An
Dept. of Electronics Engineering

<Abstract>

We proposed a hidden Markov model(HMM), which is used as a speech recognizer, for the segmentation of the continuous speech. The continuos speech is classified into two groups. The one is composed of vowel(V), consonant(C), silence(S), liquid(L) and nasal(N) and the other is composed of stop consonant(SC), non-stop

consonant(NS), silence(S), liquid(L) and nasal(N).

The trained HMM is used for the automatic phoneme segmentation and the segmented sentences are added into the DB. This process is iterated until the number of sentences of independent speaker is 120. The results of the two segmented groups are compared. The boundaries of the phonemes, which are segmented by the trained HMM, are compared with those of the phonemes which are segmented by hand. The seed-sentences used in the training of the HMM are the 30 continuously spoken sentences, which are segmented by hand.

1. 서 론

음성에서 말의 단위가 되는 음소, 음절 또는 단어 등의 경계를 알아낼 수 있다면, 음성을 연구하는 여러 분야에 상당한 도움을 줄 수 있다. 그러나 분절 작업은 단순한 작업은 아니다. 분절 작업은 많은 인력과 자본을 요구하며, 분절된 음소의 신뢰성 또한 아주 중요하다.

기존의 제안된 분절 방법에는 미리 음절 패턴을 저장하고 저장된 패턴을 입력음성과 비교하여 경계를 결정하는 템플레이트 매칭(template matching) 방법[1], 주파수와 시간 파라메터의 변화를 검출하여 분절을 시도하는 스펙트럼 천이(spectral transition) 측정 방법[2]과 다수의 파라메터를 이용하여 음소의 경계를 찾는 대분류(broad classification)[3]방법 등이 있다. 이러한 방법들을 이용할 경우 높은 신뢰성과 용이성 모두를 만족시키기는 어렵기 때문에, 자동 음성 분절에 관한 하나의 근사적 접근법으로 은닉 마르코프 모델(HMM)을 이용한 분절 방법을 제안하고자 한다.

HMM에 관한 기본적인 이론은 1960년대 말에서 1970년대 초에 걸쳐 Baum에 의해 제안되었다. 그리고 이 모델은 Barker 및 Jeliner에 의해 1970년대에 음성 인식 분야로 응용되었다.

음성 인식에서 가장 큰 어려움은 발성 과정에서 발생하는 불확실성과 불완전성 때문이다. 여러 가지 통계학적 모델을 이용하여 이러한 어려움들을 해결할 수 있고, 그 중에서도 은닉 마르코프 모델(hidden Markov model)이 가장 널리 이용되고 있다. 이것은 마르코프 체인에서 천이 파라메터들이 음성의 시간적 변동성을 나타낼 수 있고, 발성 분포를 표현하는 파라메터들은 주파수 변동성을 나타낼 수 있으며, 이러한 파라메터를 효율적으로 계산하는 여러 가지 알고리즘이 존재하기 때문이다.

수작업으로 분절된 음성의 각 분절 단위에 음소별로 레이블(label)된 소량의 씨앗 문장으로부터 HMM의 기본 모형을 발생시키고, 이를 이용하여 새로운 문장을 분절하고 분절된 결과를 다시 훈련에 사용하는 일을 반복하여 대량의 분절된 데이터 베이스를 구축하는 방법을 이용하였다.

사용된 씨앗문장과 연속음성은 TIMIT 데이터 베이스에서 제공된 61개의 음소를 5개 또는 6개의 대분류 음소로 압축하여 사용하였다.

2. 벡터 양자화

2.1. 벡터 양자화

양자화는 연속 신호를 디지털 신호로 근사화하는 과정을 말한다. 이 과정을 통하여, 데이터를 압축할 수 있고, 특히 음성 신호에서는 리던던시를 제거할 수 있다. 벡터 양자화 과정은 그림 2.1에 보인 바와 같이 벡터 양자화기가 코드북 벡터와 입력으로 받아들인 벡터들을 비교하여, 왜곡이 가장 적은 벡터의 주소를 출력으로 내보내는 과정이다. 이 과정에서 사용된 코드북 벡터는 클러스터링 알고리즘을 이용하여 교육벡터를 유사한 성질을 내포한 영역으로의 분류를 통해서 생성한다. 코드북을 생성하기 위하여 사용된 교육 벡터는 음성 신호의 시변 주파수 특성을 파라메터로 표현한 LPC 분석이나 필터 뱅크 분석의 결과 값들이 될 수 있다. 입력벡터, $\{x\} = [x_1 x_2 \cdots x_N]$를 L레벨과 N차원의 양자화기를 이용하여 벡터 공간 으로 분할할 수 있다. 분할된 각 영역의 중심벡터는 레벨이 L인 코드북의 코드 벡터에 해당한다. 이러한 코드북은 적어도 10배 이상의 교육벡터를 이용하여, 평균 왜곡이 최소가 되도록 하는 L개의 코드 벡터로 구성된다.

최적의 코드북을 생성하기 위하여 이진 분리 알고리즘이 사용되고, 초기 코드북은 Lloyd 반복법으로 구한다.

- Lloyd 반복법

① 임의의 코드북, $C_m = \{y_i\}$이 주어지면, Nearest-Neighbor 조건으로 최적의 분할 영역을 구한다.

② 중심 벡터 조건을 이용하여, 분할된 영역의 중심 세포를 구하고, 새로운 코드북 C_{m+1}을 구한다.

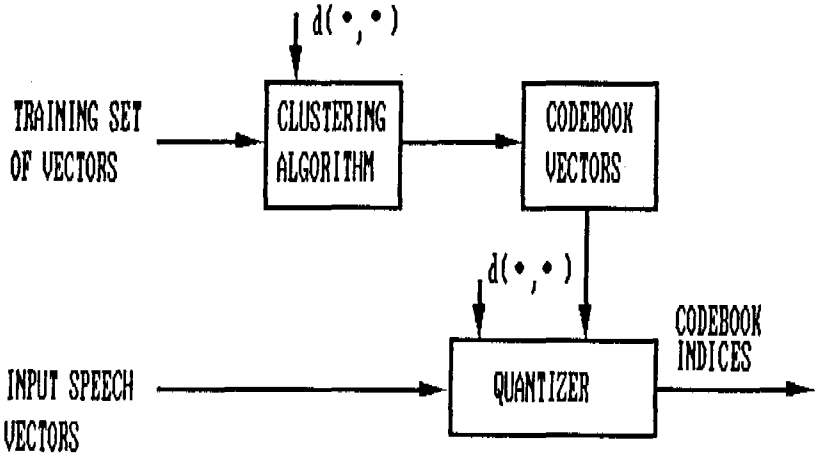

그림 2.1 벡터 양자화의 블록도

3. 은닉 마르코프 모델(Hidden Markov Model)

HMM은 관측벡터가 이산적이고 유한한 경우의 이산 관측 HMM과 연속적인 경우의 연속 관측 HMM의 두 가지로 구분된다. 연속 관측 HMM의 경우 연속 확률 밀도 함수로 보통 가우시안 분포 함수를 사용하므로 확률계산에 따른 계산량이 방대하고, 이산 관측 HMM은 전 처리기로 벡터 양자화기를 사용하므로 관측 벡터를 양자화하는 과정에서 발생하는 양자화 오차가 발생하는 단점이 있다.

3.1. 연속 은닉 마프코프 모델[4]

연속 관측 심벌 분포는 가우시안 함수를 이용하여 표현한다. 여기서 이용된 가우시안 함수는 관측의 평균 벡터와 공분산 행렬을 파라메터로 이용하며, 관측 심벌 분포는 임의로 초기화된 유한개의 혼합 계수(mixture coefficient)를 곱한 가우시안 함수의 합으로 나타낸다. 실제 관측 심벌의 분포는 가우시안 분포의 형태가 아니라, 가중치를 달리하는 가우시안 분포를 이용하여 실제 관측 심벌의 분포를 최대로 근사화한 것이다.

3.1.1. 연속관측 확률 밀도(continuous observation probability)

이산 HMM은 전처리 부분에서 양자화 과정을 거치므로 실제의 연속 관측의 중요한 특성이 손실될 수 있다. 이러한 손실을 방지하기 위하여 연속 관측 밀도를 사용한다. 확률 밀도 함수의 파라메터를 재한하여 이산 HMM에서 사용한 재추정 알고리즘을 적용한다. 사용된 모델 확률 밀도 함수의 가장 일반적인 형태는 다음과 같다.

$$b_j(o) = \sum_{k=1}^{M} c_{jk} N(o, \mu_{jk}, U_{jk}), \quad 1 \le j \le N \tag{3.1}$$

여기서,

N : 가우시안 분포
μ_{jk} : 평균 벡터
U_{jk} : 공분산 행렬

사용된 혼합 계수(mixture coefficient)에 대한 조건은 다음과 같다.

$$\sum_{k=1}^{M} c_{jk} = 1, \qquad 1 \le j \le N. \tag{3.2}$$

$$c_{jk} \ge 0, \quad 1 \le j \le N, \quad 1 \le k \le M. \tag{3.3}$$

위의 수식은 관측 벡터, 평균 벡터 그리고 공분산 행렬을 이용하여 가우시안 분포를 정의하고, 각 상태에서 혼합 계수를 이용하여 연속 관측 심벌 분포를 표현하였다. 그림 3.1은 다상태 단일 밀도 분포(multistate single-density distribution)를 나타낸 것이다. 임의의 i

상태에서 내부적으로 혼합 계수의 수에 해당하는 부(副) 상태가 가우시안 분포 형태를 이루고 있다는 것을 보여주고 있다.

식(3.4)-(3.7)은 가우시안 분포의 변수들에 대한 재추정 공식을 나타내고 있다. 프라임은 벡터의 전치를 의미하고, $\gamma_t(j, k)$는 현재의 상태가 k번째 혼합 성분을 가지고 시간 t에서 j상태에 존재할 확률이다.

$$\overline{c_{jk}} = \frac{\sum_{t=1}^{T}\gamma_t(j, k)}{\sum_{t=1}^{T}\sum_{k=1}^{M}\gamma_t(j, k)} \tag{3.4}$$

$$\overline{\mu_{jk}} = \frac{\sum_{t=1}^{T}\gamma_t(j, k)\cdot o_t}{\sum_{t=1}^{T}\gamma_t(j, k)} \tag{3.5}$$

$$\overline{U_{jk}} = \frac{\sum_{t=1}^{T}\gamma_t(j, k)\cdot (o_t-\mu_{jk})(o_t-\mu_t)'}{\sum_{t=1}^{T}\gamma_t(j, k)} \tag{3.6}$$

$$\gamma_t(j, k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right]\left[\frac{c_{jk}N(o, \mu_{jk}, U_{jk})}{\sum_{m=1}^{M}c_{jm}N(o_t, \mu_{jm}, U_{jm})}\right] \tag{3.7}$$

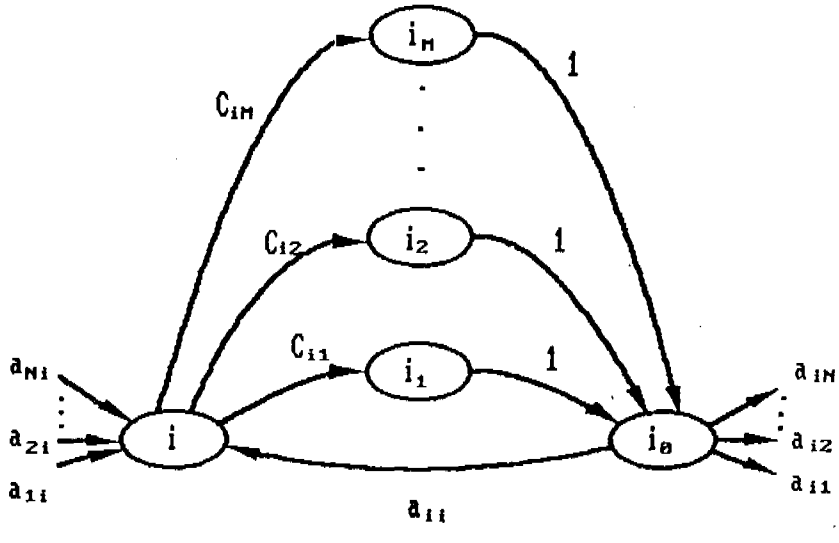

그림 3.1 혼합 밀도를 이용한 상태도

4. 실험 및 결과 분석

4.1. 은닉 마르코프 모델(HMM)의 구현

여러 가지 형태의 HMM 가운데 음성 인식에 가장 많이 사용되고 있는 좌-우 구조 모형을 사용하였다. 그림 4.1에 보인 것처럼 전체 상태는 5개이지만, 자기 상태 천이가 일어날 수 있는 실제 상태는 3개이다. 모든 상태는 최대 상태 천이를 두 개로 제한하고 있다.

상태 S1과 S5는 각각 시작과 끝 상태로 후에 독립적으로 교육된 HMM의 공통 파라메터 연결 과정에서 음소별 모델의 연결 상태로 이용된다. 이러한 모델을 구현하기 위한 상태 천이 행렬의 제약 조건과, 이를 만족하는 상태 천이 행렬을 아래에 각각 나타내었다.

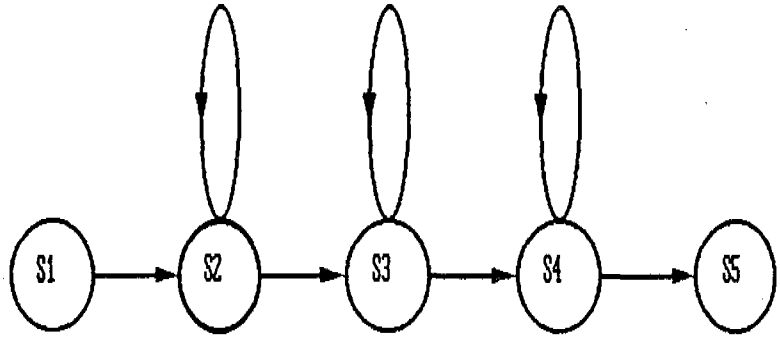

그림 4.1 사용된 HMM의 모형

1) 제약 조건

$$a_{jk} = 0, \qquad j < k, \quad i + 1 \leq j, \tag{4.1}$$

$$\sum_{j=1}^{3} a_{ij} = 1, \qquad i = 1, 2, 3. \tag{4.2}$$

2) 상태 천이 행렬

$$A = \begin{pmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \end{pmatrix}$$

구현된 시스템의 전 과정을 그림 4.2에 나타내었으며, 각 단계에서 수행하는 기능은 다음과 같다.

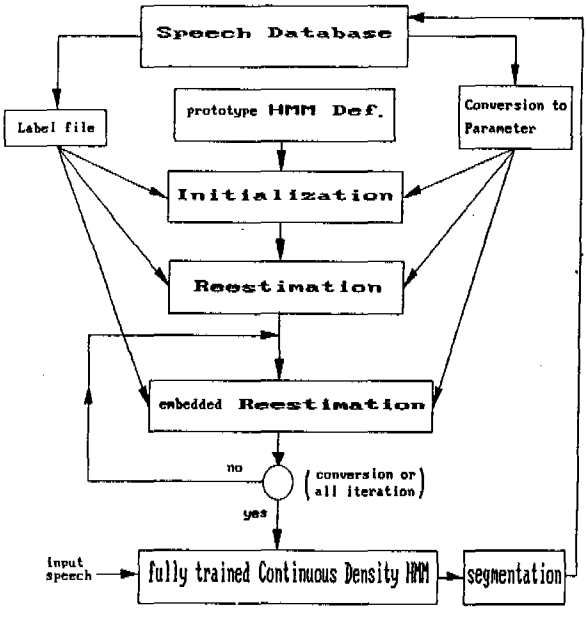

그림 4.2 HMM 교육 과정의 블록도

원형 HMM(prototype HMM) 정의 단계에서는 구성될 HMM의 파라메터 종류를 정의하게 된다. 각 상태의 data-stream의 수, 각 stream의 혼합계수 등을 설정하여 기본적인 HMM의 형태를 갖춘다.

초기화 단계에서는 데이터 베이스에서 제공된 연속음 문장을 이용하여 같은 음소별로 모델을 초기화한다. 이 단계에서는 분절과 클러스터링이 동시에 일어난다. 다중 혼합계수(multiple mixture coefficient) 모델을 구현하기 때문에 modified K-means 클러스터링 알고리즘을 초기 사이클에 이용하고, 이후의 반복적인 사이클에서는 Viterbi 정렬을 이용한다.

재추정 단계에서는 초기화 단계에서 생성된 각 음소별 모델의 파라메터를 Baum-Welch 재추정 알고리즘을 이용하여 독립적으로 재추정하여 모델을 갱신하게 된다.

모델 교육의 마지막 단계에서는 독립적으로 교육된 모델들을 동시에 갱신하여, 각 모델들의 상관성을 높이기 위해서 공통 파라메터를 연결한다.

이러한 단계를 통하여 교육된 모델을 Viterbi 알고리즘[5]을 이용하여 입력 음성을 인식하고, 데이터 베이스에 저장된 음성의 레이블 데이터와 비교하여 인식 정도를 측정한다.

4.2. 실험

본 연구에 사용된 음성 데이터 베이스는 TIMIT 데이터 베이스이며, 수작업으로 연속음

문장을 음소별로 분절하여 각 문장에 대하여 음소의 경계 정보를 저장하고 있다. 10명의 남성화자가 발성한 30개의 연속 문장을 이용하여 모델을 교육하였고, 30명의 남성화자가 발성한 90개의 연속 문장을 30개 문장씩 3개의 그룹으로 나누었다. 음소별 인식을 기초 분절된 결과를 데이터 베이스 확장에 사용하였고, 확장된 데이터 베이스를 이용하여 모델을 교육하고, 다시 새로운 문장을 분절하는 과정을 반복하였다.

사용된 음성 신호의 특징 벡터는 해밍 윈도우(Hamming window)를 이용한 멜 주파수 계수(mel-spectral coefficient)이고, 파라메터 값들은 다음과 같다.

- Window type : Hamming window
- Frame period : 10 msec
- Window duration : 25 msec
- Preemphasize Factor : 0.97
- Vector size : 12
- Mel-spectral analysis : 24

멜 필터 뱅크 분석을 위하여 사용된 멜 척도는 식(4.3)에 의해 24채널로 분석하였고, MFCC(mel-frequency cepstral coefficient)는 이산 코사인 변환을 대수 필터 뱅크 출력 m_j에 적용하여 계산하였으며, 식(4.4)와 같이 표현된다.

$$\mathrm{Mel}(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \tag{4.3}$$

$$c_i = \sum_{j=1}^{P} m_j \cos\left(\frac{\pi * i}{P}(j-0.5)\right), \quad 1 \le i \le 12, \quad P = 24. \tag{4.4}$$

실험에 사용된 음소의 종류, 분류(class), 각 분류에서 모델의 수, 그리고 각 모델당 상태의 수를 표 4.1과 표 4.2에 보였다.

이경우 HMM의 파라메터는 표 4.3와 같다. 파라메터는 혼합 계수를 3개와 4개 그리고 stream의 수를 2개와 3개로 설정하고 실험을 반복하였다.

표 4.1 사용된 음소의 종류

분류(class)	분류1(class 1)		분류2(class 2)	
음소의 종류	S	묵음	S	묵음
	L	유음	L	유음
	N	비음	N	비음
	V	모음	V	모음
	C	자음	SC	정지 자음 (stop sound)
			NC	비 정지 자음 (non-stop sound)

표 4.2 사용된 분류와 모델의 종류

분 류	분류 1	분류 2
모델수	5	6
상태수/모델	5	5

표 4.3 HMM의 파라메터

stream 정보	벡터 크기	혼합 계수의 수	파라메터의 종류
stream 1	13	3, 4	MFCC_E_D*
stream 2	13	3, 4	

* MFCC_E_D는 Mel-frequency ceptral coefficient에 에너지 계수를 첨부하여 delta 연산한 것을 의미함.

4.3. 결과

자동 분절의 기본적인 바탕은 HMM을 이용한 음성 인식기와 같은 구조를 이루고있다. 인식을 통한 자동 분절은 수작업으로 분절한 레이블 데이터의 수준까지 분절하는 것을 목표로 하고 있다. 인식율은 교육된 HMM이 분절한 결과와 레이블 데이터의 비교를 통하여 수치로 나타내었다.

그림 4.3에 데이터 양의 증가에 따른 분절율을 시험 데이터와 교육 데이터에 대하여 나타내었다. 교육 데이터 양의 증가에 따른 분절율을 나타내었으며, 사용된 교육 데이터는 자동 분절의 결과로 생성된 분절된 문장을 이용하였다. 최종적인 결과로 교육 데이터는 약 92%, 시험 데이터는 약 83%에 이르는 인식율을 보이고 있다.

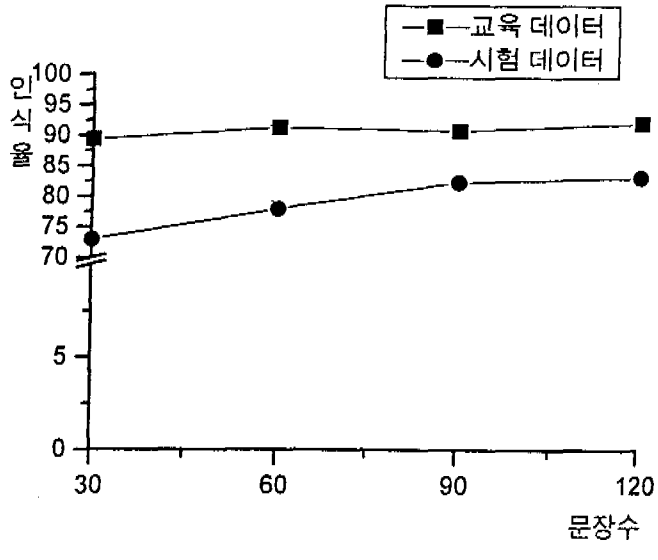

그림 4.3 교육 문장과 시험 문장의 인식율 비교

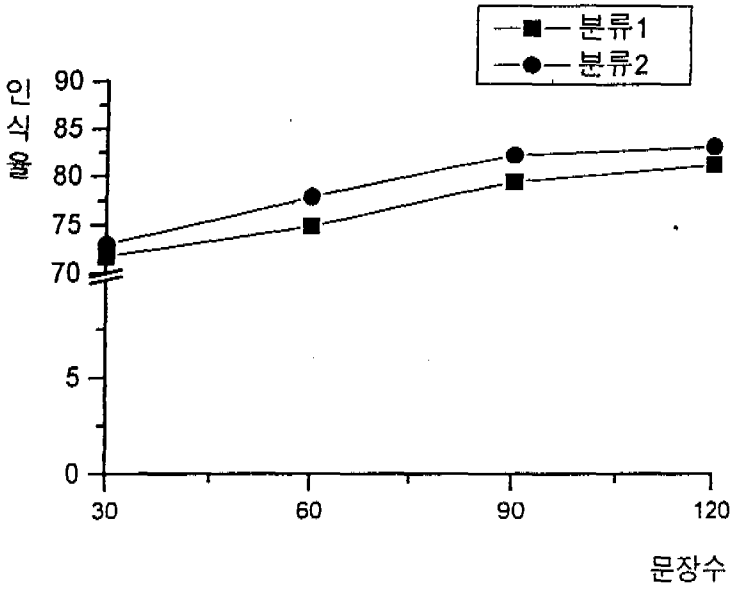

그림 4.4 분류와 인식율의 관계

음소를 5가지로 확장한 분류1의 모델과 본 논문에서 제안된 6가지로 확장된 분류2의 모델에 대하여 데이터 양을 증가시키면서 인식율을 구하였다. 분절된 결과는 그림 4.4에 보인 바와 같이 보듯이 초기의 분절결과는 거의 차이를 보이고 있지 않으나, 데이터 양과 모델의 교육횟수가 증가할수록 음소를 6가지로 확장한 분류2의 모델이 분류1의 모델보다 더 높은 인식율을 지속함으로써 다소 개선된 분절 결과를 보이고 있다.

인식된 결과는 음소를 5가지로 분류한 모델보다 본 논문에서 제안한 6가지 음소 모델에서 다소 높다. 인식율이 다소 높은 6가지 모델에 대하여 각 상태에서 혼합계수(mixture coeficient)의 수를 3개와 4개로 한 경우의 인식율을 그림 4.5에 보였다. 이 경우의 인식율은 혼합 계수의 수를 4개로 하였을 경우가 다소 높았다.

모델을 구성하는 기본 파라메터 값의 변화량을 통하여 모델의 상태를 예측할 수 있다. 모음에 대한 상태 천이 행렬의 초기값과 각 단계별 교육이 끝난 후의 상태 천이 행렬값은 다음과 같다.

1) Initial Value

0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
0.000000e+00 6.000000e-02 4.000000e-01 0.000000e+00 0.000000e+00
0.000000e+00 1.000000e+00 6.000000e-01 4.000000e-01 0.000000e+00
0.000000e+00 1.000000e+00 0.000000e+00 7.000000e-01 3.000000e-01
0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00

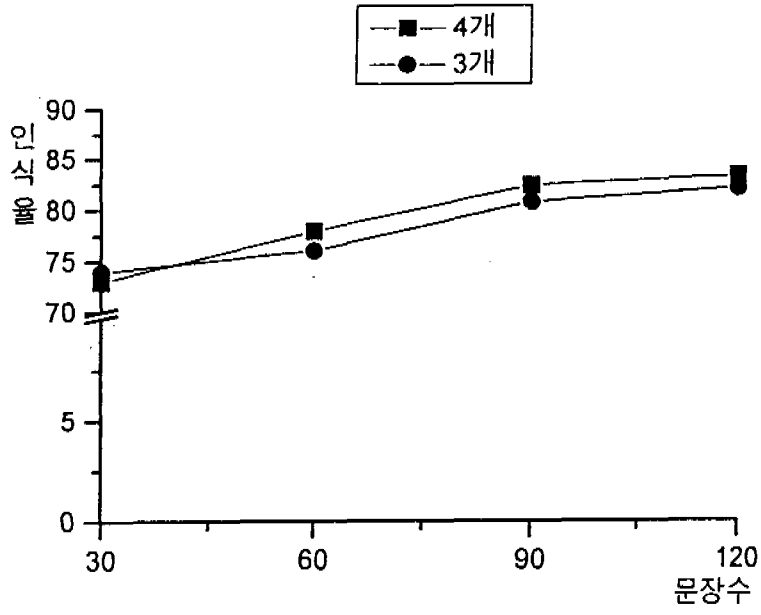

그림 4.5 혼합 계수의 수와 인식율과의 관계

2) Reestimation

0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
0.000000e+00 6.366875e-01 3.633125e-01 0.000000e+00 0.000000e+00
0.000000e+00 1.000000e+00 7.561436e-01 2.438564e-01 0.000000e+00
0.000000e+00 1.000000e+00 0.000000e+00 6.890103e-01 3.109897e-01
0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00

3) embedded Reestimation

0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
0.000000e+00 5.883878e-01 4.116121e-01 0.000000e+00 0.000000e+00
0.000000e+00 1.000000e+00 7.203436e-01 2.796564e-01 0.000000e+00
0.000000e+00 1.000000e+00 0.000000e+00 5.984709e-01 4.015292e-01
0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00

초기값은 모든 모델에 대하여 임의로 설정하였다. 두 번째 Reestimation단계가 끝난 후, 상태의 천이 가능성은 자기 상태 천이를 더욱 강화하였고, 상대적으로 다음 상태로의 천이를 어렵게 하였다. 마지막 단계가 끝난 후 상태 천이 행렬값은 다소 다른 상태로의 천이 확률을 증가시켜 모델의 성질이 두 번째 단계의 결과와는 다소 다른 성향을 보인다. 이것은 각 모델별 공통 파라메터를 연결하는 과정의 결과로 보인다. 결국 모든 상태는 다른 상태로의 천이보다는 자기 상태 천이가 지배적임을 보였다.

"She had your dark suit in greasy wash water all year"은 실험에 사용된 연속음 문장 가운데 하나이며, 시간 영역의 파형을 그림 4.6(a)에 보였다. 이 문장의 시간 파형에 대하여 TIMIT 데이터 베이스에서 제공한 수작업에 의하여 분절한 결과와 본 논문에서 제안한 HMM에 의해 자동분절을 결과를 비교하였다. 자동 분절의 결과에서 음소별 분절이 다른 음소로 분절(삽입)되거나 삭제된 경우보다는 분절된 위치가 다소 이동된 경우가 80% 이상이었다. 특히 삽입이나 삭제의 경우는 다른 음소보다 자음 음소(본 논문에서는 정지 자음(SC)과 비 정지 자음(NC))에서 가장 많이 일어났다.

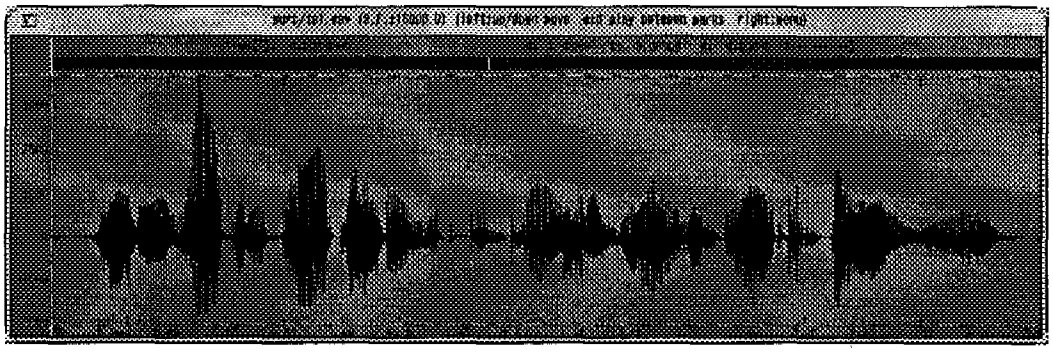

그림 4.6(a) 시험 음성의 시간 파형

1) 수작업에 의한 분절의 결과 (단위 : sec)

0.03	0.12	0.24	0.28	0.35	0.41	0.48	0.52	0.58	0.61	0.74	0.82	0.99	1.05	1.11	시 간
S	S	NC	V	V	SC	NC	L	SC	SC	L	SC	NC	V	V	분절음

1.16	1.2	1.23	1.26	1.29	1.39	1.43	1.47	1.54	1.61	1.69	1.73	1.86	1.94	1.97	2.1
SC	SC	SC	NC	N	N	SC	SC	L	V	NC	V	L	V	L	NC

2.15	2.22	2.26	2.36	2.4	2.45	2.48	2.57	2.63	2.66	2.73	2.76	2.8	2.83	2.9	3.06
L	V	SC	L	L	V	L	V	L	V	L	V	N	N	L	S

2) HMM에 의한 자동 분절 결과 (단위 : sec)

0.04	0.12	0.24	0.36	0.41	0.48	0.52	0.57	0.61	0.75	0.81	0.99	1.06	1.11	1.15	시 간
S	S	NC	V	SC	NC	V	SC	SC	L	SC	NC	V	V	SC	분절음

1.18	1.2	1.24	1.3	1.37	1.42	1.47	1.53	1.63	1.69	1.73	1.9	1.94	1.97	2.05	2.09
V	SC	SC	N	N	SC	SC	L	V	NC	V	L	V	L	NC	NC

2.17	2.22	2.25	2.36	2.48	2.59	2.63	2.66	2.72	2.81	2.84	2.89	2.92	3.06
L	V	SC	L	L	V	L	V	L	V	SC	L	SC	S

그림 4.6(b) 시험 음성의 분절 결과

5. 결 론

본 연구에서는 TIMIT 데이터 베이스에서 제공한 61개의 음소를 크게 5가지와 6가지로 압축한 음소를 이용하여 HMM을 교육하고, 이를 이용하여 화자 독립의 연속 문장을 분절하는 과정을 반복하여 데이터 베이스의 확장을 시도하였다. 교육과 시험에 사용된 연속 문장의 음성은 120개이지만, 음소별 데이터는 모음 1379개, 유음 708개, 비음 418개, 묵음 421개, 정지 자음 498개 그리고 비 정지 자음 1395개로 총 사용된 음소는 4819개이다.

사용된 모델은 음성신호 처리에 가장 많이 사용되는 좌-우 구조이며, 상태수는 5개로 제한하였다. 상태1과 상태5는 시작과 끝 상태로 실제 자기 상태 천이가 일어나지 않도록 하였다.

본 연구에서는 최상의 음소별 분절을 수작업에 의하여 분절된 음소를 목표로 하고 있다. 음소의 경계를 분석한 결과를 근거로 분절의 위치가 음소의 천이 영역 구간 내에서 분절된 경우가 많았다. 분절에 이용된 문장이 화자 독립[6] 음성임에도 불구하고 분절된 결과가 80% 이상의 인식율을 보였다.

오인식의 많은 부분이 음소의 경계부분에 존재하는 천이구간에서 발생하였다. 따라서 이러한 천이 구간 내에서의 음소 경계를 반영할 수 있는 알고리즘이 개발된다면, 음성 자동분절기로 HMM을 이용할 수 있을 것이다.

참 고 문 헌

[1] T. Svendson and F.K. Soong, "On the Automatic Segmentation of Speech Signals," Proc. ICASSP-87, Vol.1, pp.77-80, Dallas, 1897.

[2] R.A. Obrecht, "A New Statistical Approach for Automatic Segmentation of Continuous Speech Signals," IEEE Trans. on Acoustic, Speech, Signal Processing, Vol.36, No.1, January, 1988.

[3] R.A. Cole and Lily Hou, "Segmentation and Broad Classification of Continuos Speech," ICASSP-88, s10.12, New York, 1988.

[4] L.R. Rabiner, "A Tutorial on Hidden Markov Model and Applications in Speech Recognition," Proc. IEEE, Vol.77, 1989.

[5] G.D.Forney, "The Viterbi Algorithm", Proc. IEEE, Vol.61, 1973.

[6] K.F. Lee and H.W. Hon, "Speaker-Independnt Recognition Hidden Markov Models," IEEE Trans. on Acoustic, Speech and Processing, Vol.37, No11, 1989.