HMM을 이용한 연속 음성 인식의 화자적응화에 관한 연구

A Study on the Speaker Adaptation of a Continuous Speech Recognition using HMM

김 상 범*, 이 영 재**, 고 시 영***, 허 강 인*

(Sang Bum Kim*, Young Jae Lee**, Si Young Koh***, Kang In Hur*)

요 약

본 연구에서는 음절 단위의 HMM을 이용하여 발성한 문장에 대해 화자 적응화 할 수 있는 방법을 제안하였다. 문장에 대한 음절 단위의 추출은 음절HMM의 연결과 viterbi세그멘테이션으로 자동화하였고, 화자 적응화는 소량의 문장과 문장의 추가에서도 시퀀셜적으로 적응화할 수 있는 MAPE(최대 사후 확률 추정)를 이용한 학습으로 수행하였다. 신문 사설에서 취한 문장에 대하여 화자 적응화한 경우의 인식율은 71.8%로 적응화 전의 결과보다 약 37% 향상되었다.

ABSTRACT

In this study, the method of speaker adaptation for uttered sentence using syllable unit hmm is proposed. Segmentation of syllable unit for sentence is performed automatically by concatenation of syllable unit hmm and viterbi segmentation. Speaker adaptation is performed using MAPE(Maximum A Posteriori Probability Estimation) which can adapt any small amount of adaptation speech data and add one sequentially.

For newspaper editorial continuous speech, the recognition rates of adaptation of HMM was 71.8% which is approximately 37% improvement over that of unadapted HMM

Ⅰ. 서 론

최근 컴퓨터의 발달과 보급이 활발해짐에 따라 전문가뿐만 아니라 각 분야의 많은 사람들이 이용하게 되었고 지식 처리 기술이나 추론과 학습가능등이 가능한 지능화와 음성, 문자, 물체에 대한 인식 기술의 발달에 따른 휴먼화가 상호보완적으로 진전되어 가고 있다. 그 중 음성에 의한 맨-머신 인터페이스는 향후 온라인 시스템, 대화 시스템 및 자동 통역등을 구현하기 위해서 필연적이다. 이러한 시스템을 구현하기 위해서는 먼저 음성 인식 기술의 연구가 우선되어져야 한다.

현재의 음성 인식에는 DP매칭, HMM 및 신경회로망으로 처리하는 연구가 계속되고 있다. DP매칭 법은 시계열 패턴의 시간축 상에서의 비선형 신축을 허용하여 패턴을 찾는 방법으로 시간적 구조의 변동을 잘 흡수할 수 있으나 화자의 개인차 등에 기인하는 스펙트럼 그 자체의 변동에 대해서는 처리가 어렵다. HMM법은 음성의 변동을 통계적으로 처리하고 이 통계량을 확률 형태의 모델에 반영하여 음성을 인식하는 방법이며, 확률 모델을 사용하기 때문에 개인차나 조음 결합의 영향 등에 의한 음성 패턴의 변동을 반영하기 쉽고 음소나 음절 단위의 모델을 단어, 문장 등의 단위로 확장 할 수 있다. 또 신경회로망은 교사 신호들의 학습으로 원하는 입출력간의 매핑에 의해서 패턴을 분류하는 것으로 음소나 음절 인식 단계이다.

현재 연속 음성 인식을 위한 HMM의 연구는 국내외적으로 활발히 진행되고 있으며 인식율의 향상을 위한 많은 연구 보고가 있었다.[1, 2, 4]

일반적으로 음성에서 특정 화자 모델에 의한 경우가 인식율이 좋으나, 이 경우 모델을 학습하기 위해서는 특정 화자의 음성을 여러번 발성해야 하는 단점이 있다. 이러한 단점을 해결하기 위하여 화자 적응화가 필요하다.

따라서 본 연구에서는 음절 단위 HMM에 MAPE 기법

*동아대학교 전자공학과
**창신전문대학 전산정보처리과
***경북산업대학교 전자공학과

접수일자 : 1995년 9월 20일

을 이용하여 연속 음성에 대하여 화자 적응화를 수행하였다.

과거의 화자 적응화는 표준 모델의 학습과 같은 방법으로 목적으로서 추출한 음절 데이터를 이용하여 Baum-Welch 알고리즘에 의한 최우 추정법(Maximum Likelihood Estimation)으로 다수 화자 HMM의 평균 벡터만을 학습하므로 써 비교적 양호한 결과가 나왔다고 보고되어 있다.[2] ML 추정에서는 카테고리마다 모델의 학습 패턴을 다수개 준비해 놓고 그것을 학습시에 모두 주어 파라메터를 추정 갱신하고 있다. 그 때문에 문 또는 단어의 적응화용 데이터를 음절 카테고리마다 화면을 보고 추출하여야 하므로 숙련과 많은 노력이 필요하며, 학습시 데이터를 일괄해서 공급해야 하므로 추가적인 적응화는 불가능하다.

이 경우에서는 추가된 데이터와 과거 데이터를 합쳐서 처음부터 다시 추정해야 하므로 온라인 시스템에서 실시간 적용이 어렵다.

이에 반해, 본 논문에서는 발성한 문장에 대응하는 음절 라벨에 따라 만든 음절 HMM을 연결하여 문 HMM을 구성하고, 문HMM에 대응하는 발성 문장에서 viterbi 알고리즘으로 음절 경계를 추정한다. 화자 적응화는 소량의 문장과 문장의 추가에서도 적응화 할 수 있는 MAPE을 이용하였다.[6] 연속 음성 인식 실험에서는 화자 적응화 전의 방법과 비교한 결과 위 방법이 우수하였다.

Ⅱ. 지속 시간 확률 밀도 분포 HMM

본 논문에서 사용된 HMM 모델은 음성의 과도구간의 표현을 할 수 있는 지속 시간 확률 밀도 분포 HMM을 사용하였다. HMM에서 상태 천이 확률은 어느 상태가 계속 지속되어 다음 상태로 천이 할 것인가를 나타내는 확률이다. 그림 1.에서 상태 i가 n 시간 지속될 확률은

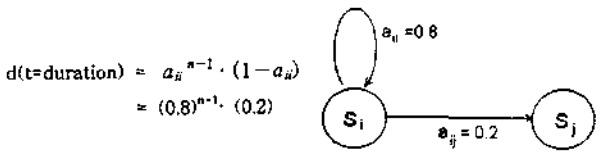


그림 1. 천이 확률과 지속 시간
Fig 1. transition probability and duration time

이 되고 n의 증가에 따라 지수 함수적으로 감소한다. HMM에서 상태가 지속되는 시간은 음성 세그맨트의 길이를 나타내며 일반적으로 음성 세그맨트의 길이는 감마 분포나 포아송 분포에 가까운 것으로 알려져 있다. 이러한 지속 시간은 음성 세그멘트의 발성 시간을 나타내는 중요한 정보이므로 인식에서 이를 고려하는 것이 바람직하다. 상태의 지속 시간 제어를 통계적으로 실시하기 위해 a_{ii}의 자기루프 천이를 제거하고 대신에 상태 i가 지속될 시간장 τ에서 지속될 지속 시간의 확률 $d_i(\tau)$을 구하여 이것을 새로운 파라메터로 추가한다. 여기서 지속 시간 확률은 $\sum d_i(\tau)=1$이다. 이 파라메터를 추가하면 연속 출력 확률 $^{\tau}$밀도 분포 HMM의 경우 Baum-Welch의 재추정 알고리즘은 다음과 같이 변화된다.

$$\alpha(i, j)=\sum_{j}\sum_{\tau\le t}\alpha(j, t-\tau)a_{ji}d_i(\tau)\prod_{n=1}^{\tau}b_{ji}(o_{t+1-n}) \quad (1)$$

$$\beta(i, j)=\sum_{j}\sum_{\tau\le T-t}a_{ij}d_j(\tau)\prod_{n=1}^{\tau}b_{ij}(o_{t+n})\beta(j, t+\tau) \quad (2)$$

여기서 식(3)을 정의하면,

$$\gamma_t(i, j, \tau)=\frac{\alpha(i, t-\tau)\,a_{ij}d_j(\tau)\prod_{n=1}^{\tau}b_{ij}(o_{t+1-n})\beta(j, t)}{P(o|M)} \quad (3)$$

천이확률 a_{ij}와 정규 분포의 파라메터 μ_{ij}, $\sum_{ij}$의 추정식은 각각 다음 식으로 주어진다.

$$a_{ij}=\frac{\sum_{t}\sum_{\tau\le t}\gamma_t(i, j, \tau)}{\sum_{t}\sum_{j}\sum_{\tau\le t}\gamma_t(i, j, \tau)} \quad (4)$$

$$\mu_{ij}=\frac{\sum_{t}\sum_{\tau\le t}\gamma_t(i, j, \tau)\sum_{n=1}^{\tau}o_{t+1-n}}{\sum_{t}\sum_{\tau\le t}\gamma_t(i, j, \tau)\tau} \quad (5)$$

$$\sum_{ij}=\frac{\sum_{t}\sum_{\tau\le t}\gamma_t(i, j, \tau)\sum_{n=1}^{\tau}(o_{t+1-n}-\mu_{ij})(o_{t+1-n}-\mu_{ij})^t}{\sum_{t}\sum_{\tau\le t}\gamma_t(i, j, \tau)\tau} \quad (6)$$

또 지속 시간 확률 $d_j(\tau)$의 추정치는

$$d_j(\tau)=\frac{\sum_{t}\sum_{i}\gamma_t(i, j, \tau)}{\sum_{t}\sum_{i}\sum_{\tau\le t}\gamma_t(i, j, \tau)} \quad (7)$$

로 된다. 식(7)의 추정만으로 학습 회수가 진행되면 지속 시간 확률 분포의 차가 너무 크게 나타나므로 식(8)과 같이 가중치 평균을 이용해서 확률 분포의 스무딩을 실시한다.

$$d_j(\tau)=\begin{cases}\{2d_j(\tau)+d_j(\tau+1)\}/3 & \text{if } \tau=1\\ \{d_j(\tau-1)+2d_j(\tau)\}/3 & \text{if } \tau=\Delta T\\ \{d_j(\tau-1)+2d_j(\tau)+d_j(\tau+1)\}/4 & else\end{cases} \quad (8)$$

Ⅲ. 화자 적응화와 최대 사후 확률 추정법에 의한 파라메터 추정

본 연구에서는 학습된 모델에 데이터를 추가하여 재학습하는 경우 기존의 데이터를 합쳐서 처음부터 다시 학습해야 하는 최우 추정법보다 우수한 최대 사후 확률 추정법에 의한 시퀀셜 연결 학습법을 이용하였다.

3.1 시퀀셜 연결 학습법

시퀀셜 연결 학습은 문에 대한 음절 HMM을 연결하여 문HMM을 만들고, 문 데이터에 대해 Viterbi 세그멘테이션을 이용하여 음절 단위로 자동 추출한 후 MAPE로 적응화시키는 방법이다.(그림2)

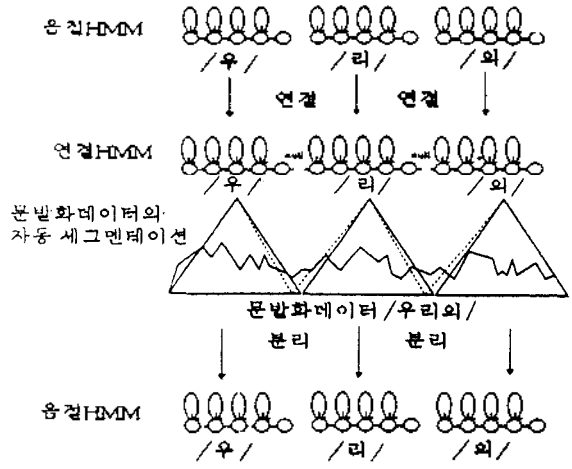


그림 2. 시퀀셜 연결 학습(발화예: 우리의)
Fig 2. sequential concatenation training(example data : woo ri ij)

3.2 최대 사후 확률 추정법

최대 사후 확률 추정법(MAP추정)은[7][8]Bayesian Successive Estimation 이라고 부르는 교사 있는 시퀀셜 학습이다. 그러므로 1개의 샘플이 주어질 때마다 사후 확률이 최대가 되도록 Θ 를 추정한다. 식(9)는 $X_1 \sim X_N$까지 N개의 샘플이 주어졌을 때의 사후 확률을 나타낸 것이다.

$$\max_{\Theta} P(\Theta \mid X_1, \cdots, X_N) =$$

$$\max_{\Theta} \frac{P(X_N \mid X_1, \cdots, X_{N-1}, \Theta)\, P(\Theta \mid X_1, \cdots, X_{N-1})}{\int P(X_N \mid X_1, \cdots, X_{N-1}, \Theta)\, P(\Theta \mid X_1, \cdots, X_{N-1})\, d\Theta} \quad (9)$$

최우 추정법에서는 식(10)과 같이 학습 샘플의 조건부 확률이 최대가 되도록 Θ를 학습하며 평균 벡터와 공분산 행렬의 추정은 각각 식(11)과 식(12)로 구한다.

$$\max_{\Theta} P(X_1, \cdots, X_N \mid \Theta) \quad (10)$$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad (11)$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu})(X_i - \hat{\mu})^T \quad (12)$$

MAP 추정을 이용해서 다차원 정규 분포의 평균 벡터와 공분산 행렬을 적응화하는 방법은 다음과 같다.

첫째, 공분산 행렬을 미리 알고 있는 경우 MAP 추정을 이용하여 다차원 정규 분포의 평균 벡터를 학습하기 위해서는 식(10)의 Θ를 식(13)으로 한다.

$$\Theta = \mu \quad (13)$$

여기서 μ는 평균 벡터이다.

그리고 1개의 샘플 X_1은 $N(\mu, \Sigma)$의 정규 분포에 따른다고 가정하고, Σ는 표준 모델의 공분산으로 미리 알고 있는 값으로 하면

$$P(X_1 \mid \mu) \cong N(\mu, \Sigma) \quad (14)$$

가 되고, 여기서 μ를 표준모델의 평균 벡터 μ_0와 공분산 행렬 K_0의 정규 분포에 따르는 사전 분포로 가정하면 식(15)가 된다.

$$P(\mu) \simeq N(\mu_0, K_0) \quad (15)$$

이상에서 정의한 확률을 적용하면 식(9)는 1개의 샘플 X1에 대하여 식(16)이 된다.

$$P(\mu \mid X_1)$$

$$= \frac{P(X_1 \mid \mu) P(\mu)}{\int P(X_1 \mid \mu) P(\mu)\, d\mu} \simeq N(\mu_1, K_1)$$

$$= C \exp\left(-\frac{1}{2}(X_1 - \mu)^T \Sigma^{-1} (X_1 - \mu)\right.$$

$$\left. -\frac{1}{2}(\mu - \mu_0)^T K_0^{-1} (\mu - \mu_0)\right) \quad (16)$$

여기서 C는 μ에 관계없는 항으로 정수이다. 추정된 평균 벡터와 공분산은 다음과 같이 된다.

$$\hat{\mu}_1 = K_0 (K_0 + \Sigma)^{-1} X_1 + \Sigma (K_0 + \Sigma)^{-1} \mu_0$$
$$\hat{K}_1 = K_0 (K_0 + \Sigma)^{-1} \Sigma \quad (17)$$

K_0는 추정 전에 가정한 공분산 행렬이다. 문헌 [7]에서는 적응화 전의 다수 화자 모델에서 각 혼합의 평균 벡터에서 구한 방법이 소개되어 있지만 공분산 행렬의 모든 값을 이용하는 경우는 어느 정도 혼합 수가 많은(적어도 10이상(≧차원수)) HMM이 아니면 정확한 K_0를 구하는 것은 불가능하다. 여기에서는 적응화 파라메터 α를 도입하고 K_0대신에 실험적으로 구한다.

$$K_0 = \alpha^{-1} \Sigma \quad (18)$$

여기서 α를 0에 근접시키면 K_0는 크게 되어 μ의 불확실성이 높고, 역으로 α를 매우 크게 하면 K_0는 작게 되어 μ의 불확실성이 낮다고 가정하는 것이 된다. 식(17)을 다시쓰면

$$\hat{\mu}_1 = \frac{\alpha \mu_0 + X_1}{\alpha + 1} \tag{19}$$

N개의 샘플을 반복해서 준 후의 추정치는 다음과 같다.

$$\begin{aligned}\hat{\mu}_N &= \frac{(\alpha + N - 1)\mu_{N-1} + X_N}{\alpha + N} \\ &= \frac{\alpha \mu_0 + \sum_{i=1}^{N} X_i}{\alpha + N}\end{aligned} \tag{20}$$

α는 모든 음절 카테고리의 각 상태에서 동일한 값으로 한다.

둘째, 평균 벡터를 미리 알고 있는 경우 N개의 샘플로 MAP추정된 공분산 행렬은 식(21)이 된다.

$$\begin{aligned}\hat{\Sigma}_N &= \frac{(\alpha + N - 1)\Sigma_{N-1} + X_N X_N^T}{\alpha + N} \\ &= \frac{\alpha \Sigma_0 + \sum_{k=1}^{N} X_k X_k^T}{\alpha + N}\end{aligned} \tag{21}$$

셋째, 평균 벡터와 공분산 행렬을 동시에 학습하는 경우는 추정해야 할 파라메터가 2개이기 때문에 사전 확률과 사후 확률은 동시 분포가 된다. 1개 및 N개의 샘플에 의해서 추정된 공분산 행렬의 추정치는 식(22)가 되고 평균 벡터의 추정치는 식(20)과 같다.

$$\begin{aligned}\hat{\Sigma}_1 &= \frac{X_1 X_1^T - (\alpha + 1)\mu_1 \mu_1^T + \beta \Sigma_0 + \alpha \mu_0 \mu_0^T}{\beta + 1} \\ \hat{\Sigma}_N &= \frac{1}{\beta + N}\left\{\sum_{i=1}^{N} X_i X_i^T - (\alpha + N)\mu_N \mu_N^T + \Sigma_0 + \alpha \mu_0 \mu_0^T\right\} \\ &= \frac{1}{\beta + N}\{X_N X_N^T - (\alpha + N)\mu_N \mu_N^T + (\beta + N - 1)\Sigma_{N-1} \\ &\quad + (\alpha + N - 1)\mu_{N-1}\mu_{N-1}^T\}\end{aligned} \tag{22}$$

Ⅳ. 인식 실험 및 결과 고찰

본 실험에서는 기존의 학습된 모델에 다른 화자의 데이터를 추가시켜 화자 적응화 실험을 하였다. 적응화 방법으로는 ML로 추정한 파라메터를 가지고 MAP 추정한 경우와 프레임을 그대로 샘플로 하는 MAP 추정한 경우에 대해 실험하였다.

4.1 분석 조건 및 음성 데이터

본 실험에 이용된 음성 데이터의 분석은 표 1과 같이 20 대 남성이 발성한 모든 음성을 10 KHz로 샘플링하여 분석창 길이 20 ms, 프레임 간격 5.0 ms의 해밍창으로 추출하고 1차 차분에 의해서 고역 강조 한 후 14차의 켑스트럼을 구하여 10차 멜 켑스트럼 계수로 변환하였다.

실험에 사용된 음성 데이터는 표 2에 나타난 바와 같이

표 1. 음성 데이터의 분석 조건
Table 1. analysis method of speech data

A/D 데이터	10 khz, 12 Bit
고역 강조	1차 차분
프레임 간격	5 ms
분석창	hamming 창
분석창 길이	20 ms
특정 파라메터	LPC Cepstrum(14차) → LPC Melcepstrum(10차)

신문 사설에서 발췌된 문장으로 구성된 10 문장의 연속 음성으로 구성되어 있다.

연속 음성은 신문 사설에서 임의로 발췌된 문장이며 따라서 각 음절의 발생 빈도가 일정하지 않고 회화 음성에 가깝게 자연스럽게 발성되었기 때문에 무음구간이 다수 포함된 음성이다. 9명의 남성 화자에 의해 5회씩 발성되었으며 그 중에서 6명 2회분을 학습용 데이터로 나머지 3명중 3회분은 적응화용 데이터로 나머지 2회분은 평가용으로 사용하였다.

표 2. 신문 사설에서 발췌된 연속 음성
Table 2. newspaper editorial continuos speech

1) 우리의 생활 문화가 문화국민의품격을 잃고 있다.
2) 말이란 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 언어 교육이 중요하다.
7) 가정에서의 언어 교육이 사회교육으로 연결된다.
8) 우리 사회의 언어 현실이라 해도 과언이 아니다.
9) 우리의 국력과 문화 수준에 맞는 언어 생활의 정화가 시급하다.
10) 방송이나 매스컴 종사자들의 엄격한 언어 통제가 필요하다.

4.2 실험 결과 및 고찰

본 논문의 실험에서 사용된 9명의 화자중(5회 발성) 3사람은 무한 반향의 방송국에서 녹음하였고 나머지 6명분은 컴퓨터 및 워크스테이션이 가동 중인 실험실에서 녹음하였다. 이중 교내 방송국에서 녹음한 3명분의 데이터와 실험실에서 녹음한 3명분의 데이터 각각의 2회발성분을 DDCHMM모델로 학습하였고, 실험실에서 녹음한 나머지 3사람분의 데이터중 3회분은 적응화용 데이터로 그중 나머지 2회분은 평가용으로 사용하였다. 이때 최대 지속 시간은 30ms로 주어 학습하였다.

기존의 학습된 모델에 적응화용 데이터로 적응화할 때 평균과 분산을 모두 적응 추정하였으며 MAP 추정에 사용되는 학습 샘플 Xi를 주는 방법에 따라 ML 추정 후의

파라메터를 MAP 추정한 경우와 viterbi 알고리즘에 의해서 추출한 프레임을 MAP 추정하는 경우에 대해 실험을 행하였다. 적응화 실험에서 최적인 적응화 계수를 구하기 위해 적응화용으로 발췌한 문장 중 2문장을 선택하여 α, β 값을 임의로 증가시켜 가면서 인식율이 높은 최적의 적응화 계수를 찾아 O(n)DP법으로 연속 음성 인식을 행하였다. 연속 음성 인식 평가에서는 세그멘테이션이 잘 되었는가가 중요하므로 음절의 대체는 고려하지 않기 위해 치환율 인식율 계산에 포함시키지 않았다.

$$\text{세그멘테이션율} = \frac{\text{입력 음절수} - \text{삽입 음절수} - \text{탈락 음절수}}{\text{입력 음절수}}$$

4.2.1 ML 추정 후의 파라메터를 MAP 추정하는 경우

ML에 의한 시퀀셜 연결 학습법은 1개의 문 데이터를 Baum-Welch 또는 Viterbi 알고리즘에 의한 학습을 실시하고 파라메터 갱신을 했다. 이때, 1문이 주어질 때마다 파라메터를 갱신하기 때문에 1개의 음절에 대응하는 프레임 수가 부족하여 불안정한 평균 벡터가 추정이 된다. 따라서 추정 전의 평균 벡터와 추정 후의 평균 벡터 사이에 선형 보간 하는 것으로 원활하게 학습된다고 생각하여 MAP 추정을 이용하였다. 그림 3에 ML 추정 후의 파라메터를 샘플로 하는 MAP 추정에 대한 블럭도이다.

그림 4는 α값 변화에 따른 인식율이며 α가 10일때 높은 인식 결과를 얻다.

4.2.2 Viterbi 알고리즘으로 추출한 프레임을 MAP 추정하는 경우

ML 추정된 평균 벡터를 이용하여 MAP 추정을 하는 경우 샘플이 되는 평균 벡터가 몇 개의 프레임에서 추정된 것인가를 알 수 없고, 프레임 수에 대한 가중치를 정확하게 줄 수 없다. 그래서 Viterbi 알고리즘에 의해서 상태마다 추출된 프레임을 그대로 MAP 추정에 사용하도록 하였다. 프레임을 그대로 MAP 추정한 경우의 블럭도를 그림 5에 나타내었다.

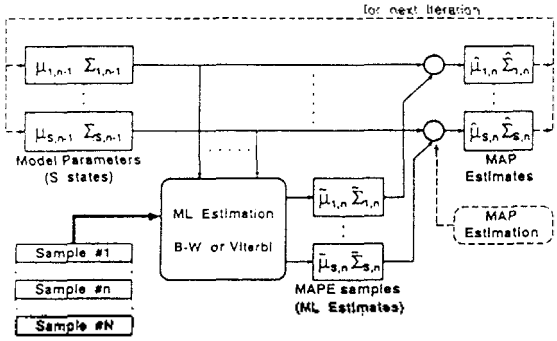


그림 3. ML 추정한 파라메터를 샘플로 하는 MAP 추정
Fig 3. MAP estimation using ML estimated parameter sample

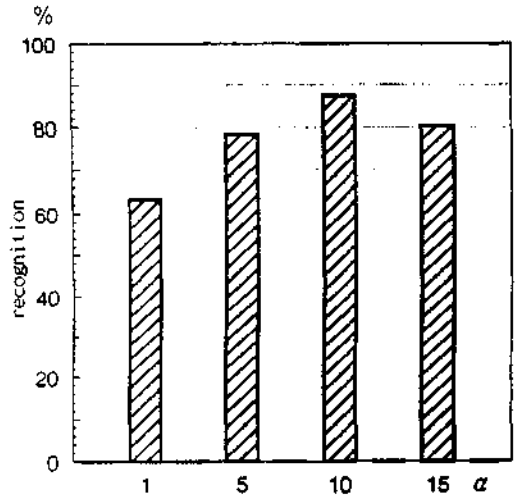


그림 4. α값 변화에 따른 인식율
Fig 4. recognition rate according to the variation of α value

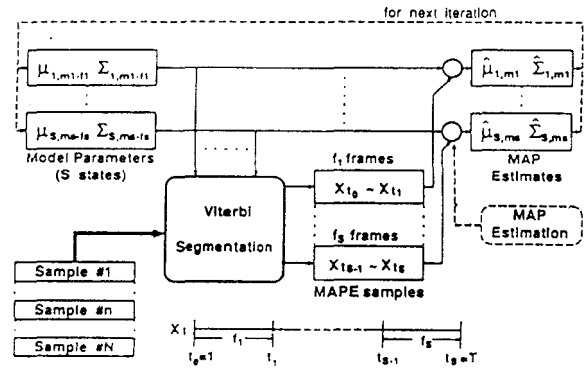


그림 5. 프레임을 샘플로 하는 경우의 MAP추정
Fig 5. MAP estimation using segmented frame sample

그림 6은 α, β 값 변화에 따른 인식율이며 $\alpha=30$, $\beta=50$일 때 가장 높은 인식율을 얻었다.

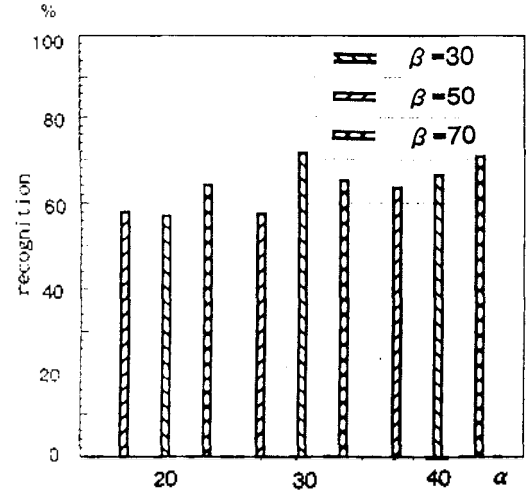


그림 6. α, β 값 변화에 따른 인식율
Fig 6. recognition rate according to the variation of α and β value

표3은 적응화 전의 인식율이며 표4, 5는 적응화 후의 인식율로서 프레임을 그대로 샘플로 하는 MAP 추정의 경우와 ML 추정 후의 파라메터를 MAP 추정하는 경우의 인식율이다.

ML 추정된 파라메터를 MAP 추정한 표5의 경우가 표4보다 인식율이 약 2% 높았으며 적응화 전의 인식율보다 약 37% 높았다.

표 3. 적응화 전의 인식율
Table 3. recognition rate of unadapted HMM

%	화자 A	화자 B	화자 C	평균
인식	22.2	43.7	38.9	34.9
치환	74	52.8	59.3	62.0
삽입	40.7	38.1	48	42.3
탈락	3.3	3.3	1.3	2.6
seg.rate	56	58.6	50.8	55.2

표 4. 프레임을 샘플로 하는 MAP추정한 경우의 인식율
Table 4. recognition rate of adapted HMM using segmented frame sample

%	화자 A	화자 B	화자 C	평균
인식	63.64	70.96	74.75	69.78
치환	35.10	28.03	23.48	28.87
삽입	41.91	39.14	32.07	37.71
탈락	1.26	1.01	1.77	1.35
seg.rate	56.81	59.85	66.16	60.94

표 5. ML추정된 파라메터를 MAP추정한 경우의 인식율
Table 5. recognition rate of adapted HMM using ML estimated parameter sample

%	화자 A	화자 B	화자 C	평균
인식	68.69	71.97	74.75	71.80
치환	27.02	24.75	22.47	24.75
삽입	18.94	18.94	16.41	18.1
탈락	4.29	6. 57	2.78	4.55
seg.rate	76.77	77.78	80.80	78.45

V. 결 론

본 연구에서는 HMM의 화자 적응화에 대해서 다음과 같은 항목과 같이 알고리즘의 개선과 검토를 실시했다.

(1) 연결 학습법에 의해 자동 세그멘테이션
(2) MAP추정에 의한 화자 적응화 실험
(3) 평균 및 분산의 적응화

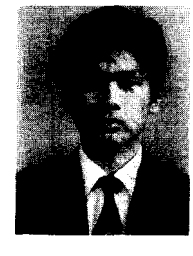
즉 음절 단위를 자동 추출한 후, 1개의 학습 샘플이 주어질 때마다 최대 사후 확률 추정법(Maximum A Posteriori probability Estimation : MAP)을 이용하므로 적은 양의 데이터로서도 적응화를 가능하게 하였다.

10문장의 연속 음성에 대해 프레임을 그대로 샘플로 하는 MAP추정과 ML추정한 파라메터를 이용하여 MAP 추정한 경우 적응 전과 비교해 최대 37%정도의 인식율 향상을 얻을 수 있었다. 그리고 인식율 향상을 위해 최적의 적응화 계수를 구하는 효율적인 연구가 해결되면 실제 온라인 시스템에 실시간 처리가 가능 할 것이다. 그리고 context HMM의 화자 적응화와 적응화 모델에 의한 화자 식별의 연구를 계속 수행할 예정이다.

參 考 文 獻

1. K-F. Lee and H-W. Hon, "Large-vocabulary speaker-independent continuous speech recognition using HMM", Proc. ICASSP, pp. 123-126, (1988)
2. 이종진, 김수훈, 허강인 "이산 지속 시간 제어 CHMM을 이용한 한국어 연속음성인식에 관한 연구", 한국음향학회지 pp81-89 (1995)
3. Hung-yan Gu, Chin-Yu Tseng, Lin-shan Lee, "Isolated-Utterance Speech Recognition Using Hidden Markov Models with Bound State Duration", IEEE Trans. Signal Processing, Vol. 39, No. 8, pp 1743-1751 (1991)
4. K. F. Lee, Automatic speech recognition : the development of the SPHINX system. Kluwer Academic Publisher, Norwell,Mass,1989
5. 中川聖一 ,平田 好充"確率出力分布型HMMの話者適應化による 日本語音節, 音節認識", 音響學會誌, Vol. 47, No 7, pp. 459-467 (1991)
6. 김상범, 이종진, 허강인, "HMM을 이용한 연속음성인식 시스템의 화자적응화에 관한 연구", 제12회 음성 통신 및 신호처리 워크샵 논문집, pp100-104 (1995)
7. Chin-Hui Lee et al., "A Study on Speaker Adaptation of the Parameters of Continuous Density Hidden Markov Models", IEEE Trans. Signal Processing, Vol. 39,No 4,pp 806-814 (1991)
8. Jean-Luc Gauvain, Chin-Hui Lee, "Bayesian Learning of Mixture Densities for Hidden Markov Models", Proc. DARPA Speech and Natural Language Workshop, Pacific Grove, pp272-277 (1991)

▲김 상 범(Sang Bum Kim) 1969년 4월 26일생



1992년 2월 : 동아대학교 전자공학과 (공학사)
1994년 2월 : 동아대학교 대학원 전자공학과(공학석사)
1994년 3월~현재 : 동아대학교 대학원 전자공학과 박사과정
※주관심분야 : 음성인식, 합성, 신경회로망

▲이 영 재(Young Jae Lee) 1956년 5월 3일

1982년 2월 : 동아대학교 전자공학과 (공학사)
1988년 2월 : 동아대학교 대학원 전자공학과(공학석사)
1996년 8월 : 동아대학교 대학원 전자공학과 (공학박사예정)
1991년 3월~현재 : 창신전문대학 전산정보처리과 조교수
※주관심분야 : 음성인식, 합성, 신경회로망, 멀티미디어

▲허 강 인(Kang In Hur)
현재 : 동아대학교 전자공학과 교수
제15권 2호 참조

▲고 시 영(Si Young Koh) 1952년 8월 16일생

1972년 2월 : 영남대학교 전자공학과 (공학사)
1983년 2월 : 영남대학교 전자공학과 대학원 (공학석사)
1992년 8월 : 동아대학교 전자공학과 대학원 (공학박사)
1972년~1979년 : 한국전자주식회사 연구개발 그룹장
1981년 : 경북개방대학 전임강사
1986년~현재 : 경북산업대학교 부교수
※주관심분야 : 음성신호처리, 생체신호처리