

HMM을 이용한 연속음성인식 시스템의 화자적응화에 관한 연구

김상범, 심장엽**, 이종진**, 허강인*
+ 동아대학교 전자공학과
** 장안전문대학 전산과

A Study on Speaker Adaptation of HMM in a Continous Speech Recognition System

"Sang Bum Kim", Jang Youp Shim, Jong Jin Lee**, Kang In Hur*
* Dept. of Electronic Eng, Dong-A Univ.
** Dept. of Computer Science, Jang Ahn Junior College

요 약

일반적으로 화자적응화는 이미 학습되어 있는 불특정 화자 모델을 표준(초기)모델로 하고 소량의 적응화용 말화로 추가적인 학습을 실시하여 특정화자 모델의 성능에 가깝게하는 기술로서 연속음성 인식에 있어서 매우 중요하다.

ML(Maximum Likelihood Estimation) 추정법을 이용한 화자적응화는 카테고리마다 모델의 학습매트릭스를 다수개 준비한 후 학습시에 일반적으로 적용시켜 모델 파라메터를 수정 갱신하므로 추가되는 화자데이터에 대해 데이터를 모두 공급하여야 한다.

본 연구에서는 문말화(文發話)데이터의 음절단위를 자동추출한 후 추가되는 화자데이터가 주어질 때마다 적응화 할 수 있는 화자적응화 방법을 검토하였다.

이 방법은 문말화 데이터를 잘라내지 않고 음절 단위를 자동추출시켜 추가 데이터마다 최대 사후확률 추정법(Maximum A Posteriori probability Estimation: MAP)을 이용하여 적응화 시키는 것으로 소수의 데이터로서도 적응화를 가능하게 하는 것이다.

본 연구에서 사용되는 음성데이터는 신문사실에서 발제한 연속음성 10문장(9사람 5번 발성)을 사용하고, 이 음성 데이터중 6명분은 HMM학습용으로 하고 나머지 3명분은 적응화용 및 평가용 데이터로 사용하였다.

6명의 화자를 DDCHMM으로 학습하고 나머지 3명분은 MAP법으로 적응화시켰다. 그 결과, 적응전과 비교해 볼때 약 32%의 인식률 향상을 얻을 수 있었다.

I. 서론

최근 컴퓨터의 발달과 보급이 눈부신게 이루어짐에 따라 일부의 전문가 뿐만 아니라 많은 사람이 컴퓨터를 이용하게 되었고 이와함께 고도의 맨-미신 인터페이스 개발의 기대가 높아지게 되었다. 음성에 의한 맨-미신 인터페이스는 인간에게 가치가 있는 일이고 음성인식 기술의 확장은 중요한 과제로 알려져 있다. 음성인식의 최종목적은 불특정 화자 연속음성 인식 이해이다. 그러나 현재 불특정화자의 제한없는 자연언어의 음성인식은 불가능한 상태이고 화자-어휘수-발성내용-발성방법 환경등에 대한 제한하에서 연구물 진행하는 경우가 많다.

음성인식의 방법으로서의 DP매칭으로 대표되는 튜플릿 매칭 방법이 화편되어 있다. DP 매칭은 시계열 패턴을 미신칭으로 압축하면서 표준 패턴과의 조화성을 실시하는 것으로 음성 패

턴의 시간축상의 변동을 흡수할 수 있고 종래의 음성인식 기술의 추위를 이루었다. 그러나 최근에 들어서 연속음성의 문장 인식이나 불특정화자의 인식등 인식 연구가 고도화함에 따라서 화음적인 인식법으로 알려져 있는 Hidden Markov Model(HMM)이 음성인식 방법의 주류중의 하나가 되었다.

특정화자 모델(이는 특정화자로부터 말화되고 학습된 모델)이 충분한 양의 데이터로 학습되어 있다면 일반적으로 불특정화자 모델(복수의 화자로 부터의 말화를 존재해서 학습한 모델)의 인식성능을 상회한다. 그러나 특정화자의 음성인식 시스템의 사용지 등록을 할 때에는 예외 대량의 학습용 데이터를 녹음할 필요가 있다. 그 때문에 화자 및 계산에 따르는 부담이 크고 시스템의 편리함을 잃어버린다. 한편, 불특정화자 모델을 이용한 시스템의 경우는 여자의 화자에 대해서도 인식을 실시할 수가 있지만 전체적으로 인식성능이 저하된다.

화자적응화는 이미 학습되어 있는 불특정 화자 모델을 표준(초기)모델로 하고 소량의 적응화용 말화로 추가적인 학습을 실시하여 특정화자모델의 성능에 가깝게하는 기술로서 연속음성 인식에 있어서 매우 중요하다.

본 연구실에서 개발되어 있는 연속음성인식 시스템에서는 음절단위로 HMM을 구성하고 있기 때문에 화자 적응화는 종래 단어 또는 문의 발생에서 목시에 의해서 추출된 음절로 실시되어 있다. 목시에 의한 추출은 속언이 필요한 작업으로 데이터가 많으면 매우 힘든 작업이다. 본 논문은 목시에 의한 추출대신에 적응화 말화의 음절 라벨계열을 주고 연결학습에 의한 문말화 그대로 음절 HMM을 적응화할 수 있다.

적응방법은 최대확률 추정법(Maximum A Posteriori Probability Estimation : MAP 추정법)^{1,2,3}을 이용하여 화자 적응화하였다.

연결학습과 MAP 추정에 의해 2,3장으로 적응화를 중단해도 그 시점에서의 최적인 파라메터가 추정되면 또 그 후에 몇회든지 추가적으로 적응화하여 파라메터의 정도를 향상시킬 수 있기 때문에 온라인 처리로 이용하는 경우에 편리하다고 생각된다.

II. 지속시간 확률밀도분포 HMM

HMM에서 상태 전이 확률은 어느 상태가 계속 지속될 것인가 다음 상태로 전이할 것인가를 나타내는 확률이다. 그림 1.에서 상태 1가 t 시간 지속될 확률은

$$d(t=duration) = a_{11}^{t-1} \cdot (a_{11}-1) \\ = (0.8)^{t-1} \cdot (0.2)$$

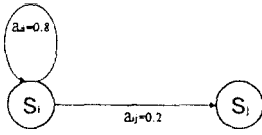


그림 1. 천이 확률과 지속시간

이 되고 1의 증가에 따라 지수함수적으로 감소한다. HMM에서 상태 지속시간은 음성 세그먼트의 길이를 나타내며 일반적으로 음성 세그먼트의 길이는 감마 분포나 포아송 분포에 가까운 것으로 알려져 있다. 따라서 기본 HMM으로는 음성파형의 과도구간이나 정상구간의 시간구조를 충분히 표현할 수 없다. 실제 각 상태에서의 지속시간은 음성 세그먼트의 발생시기를 나타내는 중요한 정보이므로 인식사에 이를 고려하는 것이 유리하다. 상태 지속시간제어를 통계적으로 실시하기 위해 a_{ii} 의 자기부피 천이를 제거하고 대신에 상태 i 와 r 시간 지속될 지속시간의 확률 $d_i(r)$ 을 구하고 이것을 새로운 파라미터로 추가한 것이 지속시간 제어 HMM이다.

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

$$\alpha(i, t) = \sum_r \sum_{i'} \alpha(i', t-r) a_{i'i} d_i(r) \prod_{n=1}^t b_{i'}(o_{t+1-n}) \quad (1)$$

$$\beta(i, t) = \sum_r \sum_{i'} a_{i'i} d_i(r) \prod_{n=1}^t b_{i'}(o_{t+1-n}) \beta(i', t+r) \quad (2)$$

여기서

$$\gamma(i, j, t) = \frac{\alpha(i, t-r) a_{ij} d_i(r) \prod_{n=1}^t b_j(o_{t+1-n}) \beta(j, t+r)}{P(o | M)} \quad (3)$$

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

$$a_{ij} = \frac{\sum_r \sum_{i'} \gamma(i', i, t)}{\sum_r \sum_{i'} \sum_{j'} \gamma(i', j, t)} \quad (4)$$

$$\mu_{ij} = \frac{\sum_r \sum_{i'} \gamma(i', i, t) \sum_{n=1}^t o_{t+1-n}}{\sum_r \sum_{i'} \sum_{j'} \gamma(i', j, t) t} \quad (5)$$

$$\Sigma_{ij} = \frac{\sum_r \sum_{i'} \gamma(i', i, t) \sum_{n=1}^t (o_{t+1-n} - \mu_{ij}) (o_{t+1-n} - \mu_{ij})^t}{\sum_r \sum_{i'} \sum_{j'} \gamma(i', j, t) t} \quad (6)$$

또 지속시간확률 $d_i(r)$ 의 추정치는

$$d_i(r) = \frac{\sum_r \sum_{i'} \gamma(i', i, t)}{\sum_r \sum_{i'} \sum_{j'} \gamma(i', j, t)} \quad (7)$$

로 된다. 추정만으로 학습회수가 진행됨에 따라서 계속시간 확률분포의 차가 너무 크게 나타나기 때문에 다음 식(8)과

같은 가중치 평균을 이용해서 확률분포의 스므딩을 실시한다.

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

III. 최대사후확률 추정법에 의한 파라미터 추정과 화자 적응화

3.1 화자 적응화

Baum-Welch 알고리즘에 의한 최우추정법(Maximum Likelihood Estimation : ML추정법)은 이용한 화자적응화는 다수화자 HMM의 평균벡터만으로 학습을 실시하여 비교적 양호한 결과를 얻었다.¹¹⁾ 그러나 이러한 ML추정법에서는 카테고리마다 모델의 학습패턴을 다수개 준비해 놓고, 학습시에 일괄해서 모델 파라미터를 추정·갱신하여야 하므로 다음 3가지 문제점이 있다.

- 1) 문발화 또는 반어발화의 적용화용 데이터들 음성 절 카테고리마다 복시에 의해서 추출할 필요가 있고 속편과 많은 시간이 요구되며
- 2) 학습시에 데이터를 일괄해서 공급해야 하므로 경우에는 추가적인 적응화가 불가능하다. 추가된 데이터와 과거 데이터가 합쳐서 처음부터 다시 추정 해야하므로 온라인 system에 적용하는 경우 효율이 나쁘고
- 3) 더욱 양호한 적응화 HMM을 추정하기 위해서는 평균벡터뿐만 아니고 공분산 행렬까지 추정하는 방법도 생각할 수 있으나 공분산 행렬의 ML 추정에는 대량의 학습샘플을 필요로 하며, 적응화 데이터를 특정화자 HMM 학습용 정도까지 대폭으로 증가 시키지 않으면 적응화는 불가능하다는 결점이 있다.

이러한 문제점을 해결하기 위하여

- 1)의 경우 문발화 데이터들 잘라내지 않고 적응화에 이용하는 방법으로써 음성단위 추출을 자동으로 실시한다. 즉 음성 절 HMM을 문발화 데이터의 음성 라벨에 따라서 연결한 HMM(이것은 다음부터 준HMM이라 부름)과 문발화 데이터 사이에서 Viterbi 세그멘테이션을 실시한 각 음절에 대응하는 프레임 구간을 자동적으로 구하는 것이다.
- 2)에 대해서 연결학습과 조합시켜 1분이 주어질 때에 파라미터를 변경하는 시퀀셜 학습을 실시한다. 본 실험에서는 1개의 학습샘플이 주어질 때마다 최적인 파라미터를 추정하는 최대 사후 확률추정법(Maximum A Posteriori probability Estimation : MAP추정법)을 HMM의 학습에 이용하였다. 또 이것은 평균 벡터의 추정뿐만 아니고 공분산행렬의 추정도 1샘플마다 실시하는 것이 가능하므로 3)에 대한 해결뿐만 아니라 HMM의 성능 향상이 기대된다.

3.2 연결 학습

연결학습은 문발화 데이터에 대응하는 음성라벨계열에 따라서 음성HMM을 연결해 문HMM을 만들고 문HMM을 문발화 데이터에 의해서 학습한다. 시퀀셜 인진학습은 1분이 주어질 때마다 파라미터 추정을 실시하는 방법이다. (그림 2참조)

3.3 최대 사후확률 추정법

최대 사후확률추정법(MAP추정)¹²⁾은 Bayesian Successive Estimation¹³⁾이라고 부르는 교차있는 시퀀셜 학습이다. 그러

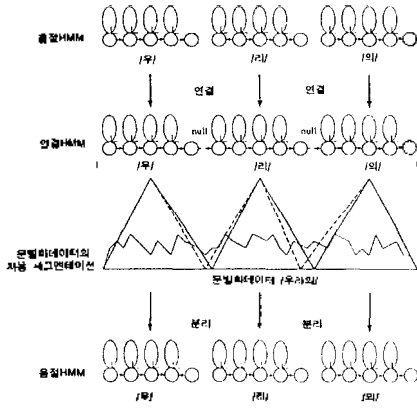


그림 2 시퀀셜 위건학습(강희애: 우리의)

므로 1개의 샘플이 주어질 때 마다 사후확률이 최대가 되도록 θ 를 추정한다. 식(9)는 $X_1 \sim X_N$ 까지 N 개의 샘플이 주어졌을 때의 사후확률을 나타낸 것이다.

$$\max_{\theta} P(\theta|X_1, \dots, X_N) = \frac{P(X_N|X_1, \dots, X_{N-1}, \theta) P(\theta|X_1, \dots, X_{N-1})}{\int P(X_N|X_1, \dots, X_{N-1}, \theta) P(\theta|X_1, \dots, X_{N-1}) d\theta} \quad (9)$$

ML의 경우 학습샘플의 조건부 확률이 최대가 되도록 θ 를 학습하여 평균벡터와 공분산 행렬이 추정된다. 이 경우 학습시에 모든 샘플을 준비할 필요가 있으며, 학습할 때에는 Baum - Welch을 이용한다. 이 추정 방법은 적응화기 전의 표준 HMM을 학습할 때에 이용한다.

$$\max_{\mu} P(X_1, \dots, X_N|\theta) \quad (10)$$

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

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

평균벡터와 공분산 행렬의 추정법에는 다음과 같은 경우가 있다.

첫째, 공분산 행렬을 미리 알고 있는 경우의 평균벡터 추정은 다음과 같다. MAP 추정을 이용하여 다차원 정규분포의 평균벡터를 학습하기 위해서는 식(10)을

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

로 한다. 여기서 μ 는 평균벡터이다.

그러나 1개의 샘플 X_1 는 $N(\mu, \Sigma)$ 의 정규분포에 따른다고 가정하고, Σ 는 표준 모델의 공분산으로 미리 알고 있는 값으로 하면

$$P(X_1|\mu) \approx N(\mu, \Sigma) \quad (14)$$

또 사전분포로서, μ 는 가장 확실한 평균벡터 μ_0 와 불확실성을 나타내는 공분산행렬 K_0 의 정규분포에 따른다고 가정하면

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

이상에서 정의한 확률을 식(10)에 대입하면

$$\begin{aligned} P(\mu|X_1) &= \frac{P(X_1|\mu)P(\mu)}{\int P(X_1|\mu)P(\mu)d\mu} \approx N(\mu, K_1) \\ &= C \exp\left(-\frac{1}{2}(X_1 - \mu)^T \Sigma^{-1}(X_1 - \mu) - \frac{1}{2}(\mu - \mu_0)^T K_0^{-1}(\mu - \mu_0)\right) \end{aligned} \quad (16)$$

식(16)을 미항하여 μ 에 관계하는 항만을 나타낸다. 여기서 C는 μ 에 관계없는 항으로 정수이다. 추정된 평균벡터와 불확실성은 다음과 같이 된다.

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

$$K_1 = K_0(K_0 + \Sigma)^{-1}\Sigma$$

K_0 는 추정전에 μ 가 어느 정도의 불확실성을 갖는가를 나타내는 공분산행렬이다. 문헌 [2]에서는 적응화전의 다수화자 모델에서 각 혼합의 평균벡터에서 구한 방법이 소개되어 있지만 전 공분산 행렬을 이용하는 경우는 어느 정도 혼합수가 많은(최어도 10이상(≥처음수)) HMM이 아니면 정확한 K_0 를 구하는 것은 불가능하다. 여기에서는 적응화 파라메터 α 를 도입하고 K_0 대신에 실험적으로 구한다.

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

여기서 α 를 0에 근접시키면 K_0 는 크게 되어 μ 의 불확실성이 높고, 역으로 α 를 매우 크게하면 K_0 는 작게 되어 μ 의 불확실성이 낮다고 가정하는 것이 된다. 식(17)에 대입하면,

$$\hat{\mu}_1 = \frac{\alpha \mu_0 + X_1}{\alpha + 1} \quad (19)$$

N 개의 샘플을 반복해서 준 후의 추정치는 다음과 같다.

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

α 는 모든 음정 카테고리 의 라 상태에 동일한 값으로 한다. 실험에 사용한 α 는 35로 하였다.

둘째, 평균벡터를 미리 알고 있는 경우의 공분산 행렬추정은 다음과 같다. N개의 샘플로 MAP추정된 공분산 행렬은 식(21)이 된다.¹⁴

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

추정에 사용되는 학습샘플 X_i 는 Viterbi 알고리즘에 의해서 추출된 프레임 집합이다. 그림3은 프레임용 샘플로 하는 MAP추정의 확률도이다.

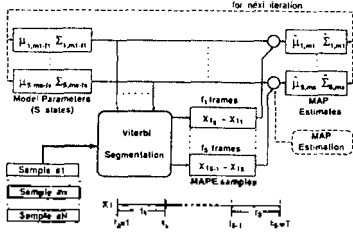


그림 3. 프레임용 샘플로 하는 MAP추정

4.3 모델의 구조와 $O(n)DP$ 법에 의한 연속음절 인식

본 실험에서 이용한 이산지속 시간장재어 연속HMM(DDCHMM)의 음절의 구조를 그림 4에 나타낸다. 각 상태의 허용 최대계속시간은 다음과 같이 설정한다.

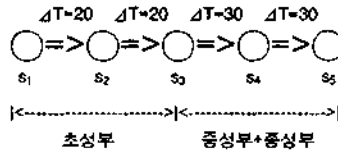


그림 4. 음절 HMM의 구조도(DDCHMM)

IV. 연속음절화확률분포 HMM의 적용과 실험

4.1 분석조건

본 실험에 이용된 음성 데이터의 분석은 표 1과 같이 20대 남성이 발성한 모든 음성을 10kHz로 샘플링하여 분석할 길이 10 ms, 프레임 간격 5 ms의 프레임으로 추출하고 1차 차분에 의해서 고역강조한 후 14차의 웨스트럼을 구하여 10차 벨 웨스트림 계수로 변환하였다.

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)

4.2 음성 자료

실험에 사용된 음성데이터는 표 2에 나타난 바와 같이 신문 사설에서 발췌된 문장으로 구성된 10 문장의 연속음성으로 구성되어 있다.

연속음성은 신문 사설에서 임의로 발췌된 문장이며 따라서 각 음절의 발생빈도가 일정하지 않고 최화음성에 가깝게 자연스럽게 발생되었기 때문에 무음구간이 다수 포함된 음성이다.

9명의 남성화자에 의해 5회씩 발생되었으며 그중에서 6명중 2회분은 학습용 데이터로 나머지 3명중 3회분은 적용화용 데이터로 2회분은 평가용으로 사용하였다.

표 2. 신문 사설에서 발췌된 연속음성

- 1) 우리의 생활문화가 문화국민의 품격을 높고 있다.
- 2) 말이란 그 나라 그 사회의 문화의 척도다.
- 3) 말이 잘 장려되고 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 통제와 필요하다.

평가용 실험은 $O(n)DP$ 법에 의한 연속음절인식 실험으로 DDCHMM을 이용하여 실시한다. $O(n)DP$ 법은 입력된 데이터에 대해서 우도가 최대가 되도록 음절HMM을 연결해 음절HMM의 카테고리를 출력하는 방법이다. 연속음절 인식이기 때문에 음절인식용의에 다음식으로 정의된 세그멘테이션율을 평가에 이용했다. 즉 연속음성인식에서 세그멘테이션이 얼마나 잘 되었는가 하는 관점에서 치환을 인식율 계산에 포함시키지 않았다.

$$\text{세그멘테이션율} = \frac{\text{입력음절수} - \text{삽입음절수} - \text{탈락음절수}}{\text{입력음절수}}$$

4.4 화자적용전과 적용후의 인식율 비교

표 3.4는 6명의 화자를 DDCHMM으로 학습한후 나머지 3사람을 MAP법으로 적용화 한 경우와 적용화 전의 인식율의 비교이다. 적용전과 비교해 불매 약 32%의 인식율 향상을 얻을수 있었다.

표3. 적용화 전의 연속음성 인식율

%	화자 A	화자 B	화자 C	평균
인식	22.2	43.7	36.9	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

표4. 적용화 후의 연속음성 인식율

%	화자 A	화자 B	화자 C	평균
인식	55.1	68.2	79.0	67.4
치환	42.9	28.8	20.5	30.7
삽입	43.9	37.6	35.1	38.9
탈락	1.3	2.3	1.8	1.8
seg.rate	54.8	60.1	63.1	59.3

V. 결론

본 연구에서는 HMM의 화자 적용화방법 및 응용에 대해서 다음과 같은 항목과 같이 알고리즘의 개선과 각종 검토를 실시했다.

- (1) 음절의 추출을 필요로 하지 않는 연결합성법에 의해 자동 세그멘테이션을 실시
- (2) MAP추정에 의한 화자적용화 실험

HMM을 이용한 연속음성인식 시스템의 화자적응화에 관한 연구

(3) 평판 및 분산의 적용화

즉 음절단위를 자동 추출한 후, 1개의 히숨션론이 주어질 때마다 최대 사후확률 추정법(Maximum A Posteriori probability Estimation: MAP)을 이용하여 적용화 시키는 것으로 적응화의 데이터로서도 적용화될 가능하게 할 수 있어 온라인 시스템에 적합하다.

향후 context HMM의 화자 적용화와 적용화 모델에 의한 화자식별에 대해 실험할 예정이다.

參 考 文 獻

- [1] 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, 1988.
- [2] Chin-Hui Lee et al., "A Study on Speaker Adaptation of the Parameters of Continuous Density Hidden Markov Models", IEE Trans. Signal Processing, Vol 39, No 4, pp. 806-814(1991)
- [3] Keinosuke Fukunage, "Introduction to Statistical Pattern Recognition", 2nd Edition, Academic Press(1990)
- [4] K-F. Lee and H-W. Hon, "Large-vocabulary speaker-independent continuous speech recognition using HMM", Proc. ICASSP, pp.123-126, 1988.
- [5] 이종진, 안정영, 허강인, "Hidden Markov Model을 이용한 한국어 연속음성 인식", 한국통신학회 부산·경남지부 학술논문 발표회 논문집, 제 1 권, pp. 72-76, 1994.
- [6] 김상범, 이종진, 허강인, "이산 시간 계층 CHMM을 이용한 한국어 연속 음성 인식에 관한 연구", 제 11회 음성 통신 및 신호처리 워크샵 논문집, 1994