1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 2(s) 호

음성인식을 위한 화자적응화 기법에 관한 연구

A Study on Methods of Speaker Adaptation for Speech Recognition

이 종 연*, 김 창 근*, 김 상 범**, 허 강 인*
(Jong Yeon Lee*, Chang Keun Kim*, Sang Beom Kim**, Kang In Hur*)

* 이 연구는 삼성자동차 과제 지원비에 의해 연구되었습니다.

요 약

본 연구에서는 음성인식을 위한 화자적응화 기법에 대해 연구하였다. 첫째로 적응화에 포함되지 않은 카테고리 음절에 대해 적응화 효과를 줄 수 있는 보간적응화 방법에 대해 연구하였다. 표준모델과 소량의 음성 데이터만으로 적응화가 가능한 MAPE(최대사후확률추정)으로 적응화한 모델의 평균벡터 변화성도를 적응화 발화에 포함되지 않은 모델에 보간적응화하는 방법이다. 둘째로 음절단위 모델을 구축한 후 적응화 하고자 하는 화자의 데이터를 연결학습법과 Viterbi 알고리즘으로 음절단위의 추출을 자동화 한 후 MAPE으로 적응화하는 방법에 대해 각각 실험을 하였다.

I 서론

화자적응화는 이미 학습되어 있는 불특정 화자 모델을 표준모델로 사용하여, 소량의 적응화용 음성을 추가적으로 학습하여 특정화자 모델에 가깝게 하는 방법으로서 연속음성 인식 시스템에 있어서 매우 중요한 연구 분야 중 하나이다. 특히 최근 컴퓨터 기술의 향상으로 인해 음성인식 시스템의 실용화가 진행되면서 주어진 화자 및 주변환경에 적응화한 수 있는 시스템 개발이 활발히 연구되고 있다. [1,2,3,4] 일반적으로 화자 적응화 방법에는 스펙트럼 특성에 기초한 방식으로 음성신호의 음향적 특징을 화자에 따라 적응화하는 방법과 기존의 학습된 파라메타를 새로 적응화하고자 환경 및 화자에 맞는 적절한 파라메타를 찾아 인식하는 방법 등이 있다. 전자의 경우는 적응시 비교적 많은 적응 데이터 및 적응 인식 시간이 소요되어지며, 후자의 경우는 음성의 변동을 통계적으로 처리하고 이 통계량을 확률형태의 모델에 반영하는 HMM에서 적응화하는 경우이다. HMM에서 Baum-Welch알고리즘에 의한 최우추정법(Maximum Likelihood Estimation)으로 적응화하는 경우는 표준모델에 적응화 할 학습 데이터를 일괄적으로 추어 학습하므로 추가적인 적응화가 불가능하며, 또 이 경우 추가된 데이터와 과거 데이터를 합쳐서 다시 추정해야 하므로 실시간 시스템에서의 적용이 어렵다. 이에 반해 최대 사후확률 추정법은 소수의 적응화 데이터로서 적응화가 가능하며, 필요시 추가적으로 적응화를 수행하여 파라메터의 정밀도를 향상시킬 수 있다. 따라서, 본 논문에서는 적응화 문장을 주어진 음절라벨에 따라 자동 추출하여 적응화하는 MAP법을 이용한 음성 인식 시스템을 개발하였으며, 소규모 테스크에서 적용가능성을 검토하였다. 또한 적응화에 포함되지 않은 카테고리 음절에 대해 적응화 효과를 줄 수 있는 보간적응화 방법에

* 동아대학교 전자공학과
** 섬유기능대학 정보기술학과

대해 연구하였다. 표준모델과 적응화한 모델의 평균 벡터 변화정도를 적응화 발화에 포함되지 않은 모델에 보간적응화 하는 방법이다. 본 논문의 구성은 2장에서 보간 적응화법을 소개하였고, 3장에서 최대사후 확률추정법에 의한 화자적응화법을 나타내었다. 4장에서는 실시간 음성인식 시스템 구현에 대해 다루었고, 5장에서는 실험 및 결과를 고찰하였고, 제 6장에서는 결론과 향후 연구과제를 제시하였다.

Ⅱ 보간적응화법

MAPE에 의한 화자적응화의 경우 적응하고자 하는 음절데이터에 해당하는 HMM 모델만 적응화 할 수 있다. 이에 적응화 데이타에 포함되지 않은 카테고리의 음절에 대해 적응화 효과를 줄 수 있는 방법으로 연결 학습에 의한 평균벡터의 학습과 스펙트럼 사상을 이용해서 비적응 혼합분포 HMM을 적응화하는 방법 등이 연구되고 있다. 따라서 본 논문에는 표준모델과 적응모델의 평균벡터의 변화정도를 적응화 발화에 포함되지 않은 모델에 보간 적응화하는 방법에 대해 실험하였다. 적응화 발화에 포함되지 않았던 카테고리 음절을 적응화 하기 위해 본 실험에 사용한 5상태 4출력의 HMM 모델의 전반부(상태 1->2->3)는 자음으로, 후반부(상태 3->4->5)는 모음에 내응하여, 화자 적응화 된 음절HMM과 적응화 전 표준모델에서의 평균 벡터의 변화 정도를 가지고 적응화 발화에 포함되지 않았던 카테고리의 음절 HMM의 모음부, 자음부를 각각 보간해서 적응화하였다. 그림 1에 보간 적응화의 예를 나타내었다. 이에 대한 보간적응화 모델을 구하는 식은 각각 식1, 2에 나타내었다.

$$\gamma_s = \frac{\sum_{k=1}^{K}\left(W_k \frac{\overline{\mu_{ks}}}{\mu_{ks}} \right)}{\sum_{k=1}^{K} W_k} \qquad (1)$$

$$\overline{\mu_{ns}}' = \gamma_s \ \mu_{ns}' \qquad (2)$$

여기서 적응화 하고자 하는 모음 또는 자음 카테고리를 C(예를 들면 /가/,/나/ 의 음절 카테고리는 다르지만, 모음 카테고리 /아/와 같은 모델), HMM의 상태를 s로 가정하였다. 화자적응화된 C를 포함한 음절 카테고리의 HMM모델이 K개($1 \le k \le K$)라 하고, 그 화자 적응화 전의 평균벡터를 μ_{ks} , 화자적응화 후의 평균벡터를 $\overline{\mu_{ks}}$로 나타내었다. 보간 적응화 하고자 하는 C를 포함한 음절 카테고리의 HMM 모델이 N개($1 \le n \le N$)라 하고, 보간 적응화 하고자 하는 평균벡터를 μ_{ns}' 로 나타내었다. 그리고 μ_{ks} 를 화자 적응화 할 때에 이용한 샘플수를 W_k로 하였다. 화자적응화는 ML추정된 평균벡터를 샘플로 하는 MAP 추정으로 실시하기 때문에 W_k는 C를 포함하는 음절 카테고리의 음절수로 하였다. 화자적응화에 의해서 μ_{ks}가 $\overline{\mu_{ks}}$로 변화한 비율 γ_s에 의해 적응화 되어있지 않은 음절 HMM의 평균 벡터를 모두 n, s(모음:1 ~3, 자음:3 ~5)에 대해서 보간하여 보간적응화 모델을 구하였다.

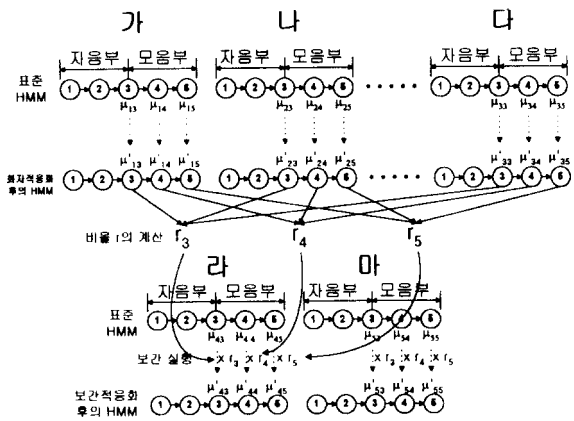

그림1. 보간적응화
Fig1. interpolation adaptation.

Ⅲ. 화자적응화와 최대사후확률 추정법에 의한 파라메터 추정

3.1 화자적응화

최우추정법(ML)에 의한 화자적응법은 많은 양의 적응화 데이터가 필요하며, 추가적 적응화의 경우 과거의 데이터와 함께 일괄해서 학습하여야 하므로 온라인 시스템 사용의 경우 비효율적이다. 이에 반해 최대사후확률 추정법(MAP)은 소수의 적응 데이터만으로 적응화가 가능하며, 시퀀셜 학습에 의해 추가적인 적응화가 가능하다. 본 실험에서는 1개의 학습 샘플이 주어질 때마다 최적인 파라메터를 추정하는 최대 사후 확률추정법(Maximum A Posteriori probability Estimation : MAP추정법)을 HMM 학습에 이용하였다. 이는 평균 벡터의 추정뿐만 아니라 공분산 행렬의 추정시 1샘플만으로 파라메타 추정이 가능한 시퀀셜 학습으로서 추가적인 적응화가 가능하다. 또한 음절 HMM모델을 발성문 데이터의 음절 라벨에 따라서 연결한 HMM 모델과 발성문 데이터와의 Viterbi 세그멘테이션에 의해 각 음절에 대응하는 프레임 구간을 자동적으로 구한 후 MAP 추정하는 인결학습법을 이용하여 적응화 실험을 행하였다.

3.2 최대 사후확률 추정법

최대 사후확률 추정법(MAP추정)은 Bayesian Successive Estimation이라고 부르는 교사있는 서켄셜 학습이다. 그러므로 1개의 샘플이 주어질 때마다 사후확률이 최대가 되도록 Θ를 추정한다. 다음의 식은 $X_1 \sim X_N$까지 N 개의 샘플이 주어졌을 때의 사후확률을 나타낸 것이다.[5,6]

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

$$\max_{\Theta} \frac{P(X_N|X_1,\cdots,X_{N-1},\Theta)\ P(\Theta|X_1,\cdots,X_{N-1})}{\int P(X_N|X_1,\cdots,X_{N-1},\Theta)\ P(\Theta|X_1,\cdots,X_{N-1})d\Theta} \tag{3}$$

평균벡터와 공분산 행렬의 추정법에는 두 가지 경우가 있다.

첫째, 공분산 행렬을 미리 알고 있는 경우의 평균벡터 학습은 다음과 같다

$$\begin{aligned}\hat{\mu}_N &= \frac{(\alpha+N-1)\mu_{N-1}+X_N}{\alpha+N} \\ &= \frac{\alpha\mu_0+\sum_{i=1}^{N}X_i}{\alpha+N}\end{aligned} \tag{4}$$

α는 모든 음절 카테고리의 각 상태에서 동일한 값으로 한다. 실험에 사용한 α는 10으로 하였다.

둘째, 평균벡터를 미리 알고 있는 경우의 공분산 행렬학습은 다음과 같다.

N개의 샘플로 MAP 추정된 공분산 행렬은 다음과 같이된다.

$$\begin{aligned}\hat{\Sigma}_N &= \frac{(\alpha+N-1)\Sigma_{N-1}+X_NX_N^T}{\alpha+N} \\ &= \frac{\alpha\Sigma_0+\sum_{k=1}^{N}X_kX_k^T}{\alpha+N}\end{aligned} \tag{5}$$

셋째, 평균벡터와 공분산 행렬의 동시 학습의 경우는 추정해야 할 파라메터가 2개이기 때문에 사전분포와 사후확률은 동시분포가 된다. 1개 및 N개의 샘플에 의해서 추정된 공분산행렬의 추정치는 다음 식이 되고 평균벡터의 추정치는 식(4)와 같다.

$$\begin{aligned}\Sigma_1 &= X_1X_1^T-(\alpha+1)\mu_1\mu_1^T+\beta\Sigma_0+\alpha\mu_0\mu_0^T \\ \Sigma_N &= \frac{1}{\beta+N}\left\{\sum_{i=1}^{N}X_iX_i^T-(\alpha+N)\mu_N\mu_N^T+\Sigma_0+\alpha\mu_0\mu_0^T\right\} \\ &= \frac{1}{\beta+N}\{X_NX_N^T-(\alpha+N)\mu_N\mu_N^T+(\beta+N-1)\Sigma_{N-1} \\ &\quad +(\alpha+N-1)\mu_{N-1}\mu_{N-1}^T\}\end{aligned} \tag{6}$$

3.3 ML추정 후 파라메터를 MAP추정.

ML에 의한 서켄셜 연결 학습법은 적응화용 데이터를 Baum-Welch또는 Viterbi알고리즘으로 추정 학습한 후 평균벡터와 분산을 MAP 추정한 경우이다. 이때, 1문이 주어질 때마다 파라메터를 생신하기 때문에 1개의 음절에 대응하는 프레임 수가 부족하여 불안정한 평균 벡터가 추정이 된다. 따라서 추정 전의 평균 벡터와 추정 후의 평균 벡터를 MAP 추정으로 신형보간하여 학습한다. 본 실험에서는 Viterbi알고리즘에 의해 상태에 대응하는 프레임을 구한 후 ML 추정하였다. 하나의 샘플로 HMM 각 상태의 평균벡터를 구할 수 있으며, 웨이트 값 α는 10일 때 최적이었다. MAP를 이용한 화자적응화 과정을 그림 2에 나타내었다.

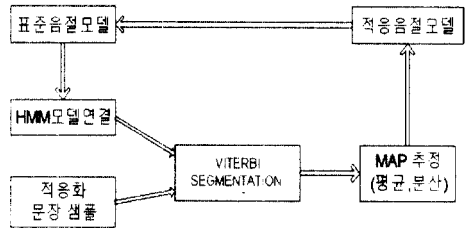

그림 2. MAP를 이용한 화자적응화 과정
Fig2. Process of speaker adaptation using MAP

3.4 Viterbi 알고리즘으로 추출한 프레임을 MAP 추정.

ML 추정된 평균 벡터를 이용하여 MAP 추정을 하는 경우 샘플이 되는 평균벡터가 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.

Ⅳ 실시간 연속음성 인식 시스템 구현

4.1 시스템 구성 개요

본 논문에서 구성한 실시간 연속음성 인식 시스템의 블럭도를 그림 3에 보였다.

그림 3에서 마이크로 입력되어 A/D 변환된 음성은 환상형 입력용 버퍼에 순환적으로 저장되도록 구성하였고, 저장된 음성에 대해 시작점과 끝점을 검출하여 무음구간을 제거한 음성 부분만을 다시 저장용 버퍼에 저장하도록 하였다.

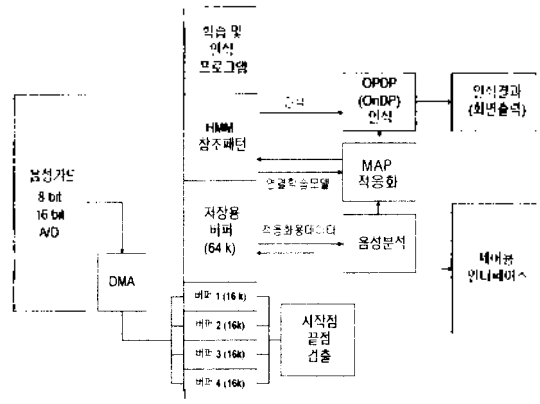

그림 3. 연속음성 인식 시스템도.
Fig3. Schematic diagram of continuous speech recognition system

저장용 버퍼에 저장된 음성을 음성분석 과정에서 10차 멜 캡스트럼을 구하여 이들 파라메터를 입력용으로 사용한 버퍼에 저장하여 테스트용 음성 파라메터로 사용할 수 있도록 하였다.
인식과정은 학습되어 구성된 HMM 참조패턴들과 입력용 버퍼에 저장된 테스트용 파라메터를 인식 알고리듬으로 패턴 비교하여 인식 결과를 PC 모니터에 문자로 출력함과 동시에 제어용 인터페이스로 제어 데이터를 보낼 수 있도록 구성하였다.

4.2 실시간 시작점·끝점 검출 방법

실시간 연속음성 인식 시스템에 입력되는 연속음성에서 음성이 시작되기 전의 무음 구간과 음성이 끝나고 지속되는 무음 구간을 제거하여 실제 연속음성 부분만 버퍼에 저장하기 위해서는 시작점과 끝점 검출이 필요하다. 본 논문에서 제시한 그림 4의 실시간 시작점·끝점 검출 시스템에서는 샘플링된 입력 음성을 DMA법으로 환상형 버퍼에 순환적으로 저장하였다.

이 환상형 버퍼는 4개의 영역으로 나누어져 있으며 음성이 버퍼에 저장되고 있는 동안 4개의 영역 중에서 첫번째 영역 버퍼1이 채워졌는가를 검사하며 채워졌으면 버퍼1의 음성 데이터(PCM)를 64개의 프레임으로 나누고 프레임별로 에너지 계산을 하며 계산된 에너지 값이 문턱값 이하이면 무음으로 간주하고, 무음이 아니면 실제 음성 저장용 버퍼에 보관한다. 이 때 음성은 버퍼2에 계속 저장되고 있다. 그리고 계속해서 다음 프레임에 대해서도 같은 방법으로 진행하여 64 프레임분(1개 버퍼영역)이 끝나면 다음 버퍼 영역에 대해서도 같은 방법으로 버퍼가 채워졌는가를 검사하고 채워졌으면 이상과 같은 방법으로 진행해 나간다.

여기서 최초로 유음(에너지 값이 문턱값 이상:시작점)이 발견되면 저장용 버퍼에 저장함과 함께 끝점을 검출하기 시작한다. 끝점 검출은 시작점 검출과 마찬가지로 에너지 계산을 하며, 시작점이 검출된 후 다시 에너지 값이 문턱값 이하인 프레임의 개수를 체크해 가면서 문턱값 이하인 프레임(무음 프레임)이 24 프레임 이상 연속적으로 지속되면 끝점으로 간주하여 한 문장의 연속음성 입력이 완료된 것으로 간주하여 입력을 종료하고 음성분석 단계로 넘어간다. 여기서 24 프레임 미만인 경우는 연속음성 중간의 묵음 구간으로 간주하고 이상의 과정을 반복한다.

음성 데이터는 입력용 버퍼에 DMA 장치에 의해 CPU와는 독립적으로 입력된다(버퍼 1 → 버퍼 2 → 버퍼 3 → 버퍼 4 → 버퍼 1).
음성이 입력되고 있는 동안에 버퍼1이 채워졌는가를 검사한다(DMA 어드레스 레지스터를 검사하면 알수 있다). 그리고 각 입력용 버퍼 영역은 다시 그림 3과 같이 64개의 프레임(256 샘플 분)으로 나눈다.

$$P(i)=\sum_{n=1}^{256} S(i)^2 \qquad (7)$$

계산된 에너지 $P(i)$가 문턱값 512 보다 작으면 무음 프레임으로 인지하여 저장하지 않도록 하였다.

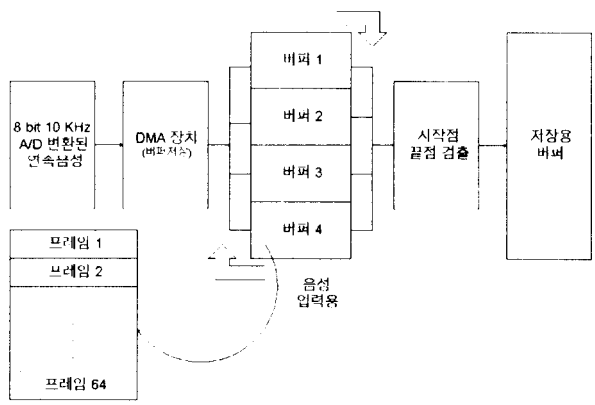

그림 4. 시작점 끝점 검출 시스템
Fig4. Schematic diagram of start and end point detection system

V. 인식 실험 및 결과 고찰

인식실험은 MAPE에 의한 화자적응화 실험과 보간적응화 실험에 대해 각각 행하였다. 적응화 방법으로는 ML로 추정한 파라메터를 가지고 MAP 추정한 경우와 프레임을 그대로 샘플로 하는 MAP 추정의 경우에 대해 각각 실현하였다.

5.1 분석 조건 및 음성 데이터

본 실험에 이용된 음성 데이터(표 2)의 분석은 표 1과 같이 20대 남성이 발성한 모든 음성을 10KHz로 샘플링하여 분석창 길이 20 ms, 프레임 간격 5.0 ms의 해밍창으로 추출하고 1차 차분에 의해서 고역 강조한 후 14차의 켑스트럼을 구하여 10차 멜켑스트럼계수로 변환하였다.

5.2 MAPE을 이용한 화자 적응화 실험

본 논문의 실험에서 사용된 8명의 화자중(5회 발성) 5사람은 잡음이 없는 곳에서 녹음하였고 나머지 3명분은 컴퓨터 및 워크스테이션이 가동 중인 실험실에서 녹음하였다. 이 중 5명분의 데이터 각각의 5회 발성분을 CHMM모델로 학습하였고, 실험실에서는 녹음한 3사람분의 데이터 중 3회분은 적응화용 데이타로, 그 중 나머지 2회분은 평가용으로 사용하였다.

표 1. 음성 데이터의 분석 조건

Table 1. analysis method of speech data

A/D 데이터	10kHz, 16it
고역강조	1차 차분
프레임 간격	5ms
분석창	Hamming 창
분석창 길이	10ms
특징 파라미터	LPC Cepstrum(14차) -> LPC Melcepstrum(10차)

5상태 4출력 left-to-right형 CHMM으로 표준 모델을 작성한 후 적응화 하고자 하는 화자의 데이터를 가지고 적응실험을 하였다. 이때 추정 파라메터로서는 평균과 분산을 각각 추정하는 경우와 평균, 분산을 동시에 MAP 추정하는 경우에 대해 O(n)DP 법으로 인식실험을 하였다. 또한 파라메터 추정방법에 따른 ML 추정 후 MAP 추정하는 경우와 Viterbi 알고리즘으로 추출한 프레임을 MAP 추정한 경우, 각각에 대해 인식실험을 행하였다. 적응화 실험에서 최적인 적응화 계수를 구하기 위해 적응화용으로 발성한 문장 중 2문장을 선택하여 α, β값을 임의로 증가시켜 가면서 인식률이 높은 최적의 적응화 계수를 찾아 O(n)DP 법으로 연속 음성 인식을 행하였다. 연속 음성 인식 평가에서는 세그멘테이션이 잘 되었는가가 중요하므로 음절의 대체는 고려하지 않기 위해 치환을 인식 계산에 포함시키지 않았다.

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

표 2. 자동차 제어문장

Table 2. Speech of car control

에어콘 켜, 에어콘 꺼
라디오 켜, 라디오 꺼
히터 켜, 히터 꺼
에어콘 강도 조정, 히터 강도 조정
라디오 볼륨 조정, 라디오 채널 조정
앞 우측 창문 열어, 앞 우측 창문 닫아
앞 좌측 창문 열어, 앞 좌측 창문 닫아
뒤 우측 창문 열어, 뒤 우측 창문 닫아
뒤 좌측 창문 열어, 뒤 좌측 창문 닫아
전화 걸기, 전화 끄기, 전화 재발신

표 3은 적응화 전의 인식률과 적응화 후의 인식률을 비교한 것으로서 상당한 인식률 향상을 얻을 수 있었다. ML 추정 후 MAP 추정하는 실험에서는 평균, 분산 추정시 시퀀셜 하게 한 문장이 추이될 때마다 MAP 추정하도록 하였고, 적응 웨이트 값은 음절별로 일률적으로 10을 주었다. Viterbi로 추출한 프레임을 MAP 추정한 경우는 문장 전체를 일괄적으로 MAP 추정하였으며, 웨이트 값은 데이터 프레임을 상태별로 세그먼테이션한 후 각 상태별로 주었다 여기서 웨이트값 α, β를 각각 30, 50으로 주어 학습하였다. 실험결과 Viterbi로 추출한 프레임을 평균과 분산을 함께 MAP 추정한 경우 추정전에 비해 77.18%의 인식률을 얻을 수 있었다.

표 3. O(n)DP 법을 이용한 화자 적응화 인식결과 (CHMM의 경우)

Table3. recognition rate of speaker adaptation using O(n)DP. (In case of CHMM)

%	인식	치환	삽입	탈락	seg. rate
적응전 인식결과	71.17	27.48	31.08	1.35	67.57
①적응후인식결과 (ML추정후-MAP) 평균추정	75.38	23.57	26.58	1.05	72.37
②적응후인식결과 (ML추정후-MAP) 분산추정	74.48	24.92	35.89	0.60	63.51
③적응후인식결과 (ML추정후-MAP) 평균+분산추정	76.43	22.82	27.03	0.75	72.22
④적응후인식결과 (FRAME-MAP) 평균추정	75.53	23.42	26.28	1.05	72.67
⑤적응후인식결과 (FRAME-MAP) 분산추정	77.18	21.77	27.48	1.05	71.47
⑥적응후인식결과 (FRAME-MAP) 평균+분산추정	77.18	22.07	29.58	0.75	69.67

5.3 보간 적응화 실험

보간 적응화 실험에 사용된 데이터는, 연속음성에서 유사한 음절을 추출하여 실험하기에는 많은 연속음성 데이터가 필요 되어지는 까닭에, 본 실험에서 만든 50음절에 대해 단음절 인식을 행하였다. 실험방법은 먼저 표준데이터와 적응화 하고자 하는 화자의 음절을 3회 발성한 데이터를 각각 화자적응화 시켰다.(/가/,/나/,/다/,/라/음절 모델)

표 4. 보간적응화를 이용한 화자 적응화 인식결과
Table4. recognition rate of speaker adaptation using interpolation adaptation.

%	적응전 인식결과	적응후인식결과 (FRAME-MAP) 평균+분산추정	보간적응화후 인식결과
인식	24.14	76.62	40.67

그리고 적응화에 포함되지 않은 모델의 평균벡타에 보간계수를 곱하여 보간적응화 시켰다.(/바/,/바/,/사/,/자/음절 모델).

괄호 안에 있는 음절의 보간적응화는 모음 "아"에 대해 보간 적응화한 예이다. 이들 모델을 가지고 적응화전의 경우, 화자적응화 한 후의 경우, 및 일부는 화자적응화 하고 나머지는 보간적응화 한 경우에 대해 각각 인식실험을 비교하였다.(표5.) 인식 결과, 보간 적응화의 경우 MAPE 적응화 보다 좋지 못한 결과를 얻었다. 이는 본실험에 사용된 C-V모델의 음소 세그멘터의 가정문제와 이를 보간하는 파라메터가 평균벡터만 한정된 관계로 인식률이 좋지 않았다. 따라서 보간 파라메터의 개선과 C-V-C모델에서의 음소 세그멘터 문제를 해결한다면, 대용량 어휘의 경우 특정 음절들에 대해 적응화 데이터를 따로 추출할 필요 없이 적응화할 수 있는 방법으로서 개선된 적응화 방법이라 할 수 있겠다.

VI. 결 론

본 연구에서는 CHMM을 음절 모델로하여 시켄셜적으로 주어진 문장을 화자 적응화할 수 있는 알고리즘의 개선과 검토를 연구하였고, 적응화에 포함되지 않은 카테고리 음절에 대해 적응화 효과를 줄 수 있는 보간적응화 방법에 대해 각각 연구하였다. 음절단위를 자동 추출한 후, 1개의 학습 샘플이 주어질 때마다 최대 사후 확률 추정법을 이용하므로 적은 양의 데이터로서도 적응화를 가능하게 하였다.

CHMM모델을 음절단위로 학습한 후 주어진 연속음성과 적응화를 실시하여 인식 실험하였다. O(n)DP법에 의한 인식 결과, 적응화한 경우의 인식률이 77.18%로 적응화 전보다 약 6% 향상되었다.

위의 연구결과 적은 데이터를 가지고 적응화하여 상당한 인식률 향상을 얻을 수 있었으며, 보간적응화의 경우는 특정 음절들에 대해 적응화 데이터를 따로 추출하여 적응화 할 필요가 없는 방법이지만 보간 파라메터 개선과 C-V-C 모델에서 세그멘터 방법을 해결해야 하는 숙제가 남아있다.

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-vocablary Speaker-Independent Continuous Speech Recognition using HMM: The SPHINX Sysytem", Proc. ICASSP, PP.123-126, 1988.

[3] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models", IEEE ASSP magazine,pp.4-17,January 1986.

[4] Chin-Hui Lee et al.,"A Study on Speaker Adaptation of theParameters of Continuous Density Hidden Markov Models", IEEETrans.SignalProcessing,Vol.39,No4, pp.806-814(1991)

[5] 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)

[6] 中川聖一, 下田好充 "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)

[7] 김상범, 이영재, 고시영, 허강인,"HMM을 이용한 연속 음성 인식의 화자적응화에 관한 연구", 한국음향학회지, pp5-11, 1995