화자적응화 연속음성 인식 시스템의 구현에 관한 연구

A Study on Realization of Continuous Speech Recognition System of Speaker Adaptation

김 상 범**, 김 수 훈*, 허 강 인*, 고 시 영***
(Sang Bern Kim**, Soo Hoon Kim*, Kang In Hur*, Si Young Ko***)

※ 이 연구는 한국과학재단의 '96 핵심전문연구 과제 지원비(KOSEF961-0921-113-1)에 의해 연구되었습니다.

요 약

본 연구에서는 소량의 음성 데이터만으로 적응화가 가능한 MAPE(최대사후확률추정)을 이용한 연속음성 인식시스템 개발에 대해 연구하였다. 음절단위 모델을 구축한 후 적응화 하고자 하는 화자의 데이터를 연결학습법과 Viterbi 알고리즘으로 음절단위의 추출을 자동화 한 후 MAPE로 적응화하였다. 자동차 제어문에 대해 화자 적응화한 경우의 인식률(O(n)DP인 경우)은 77.18%로 적응화 전의 결과보다 약 6%향상되었다.

ABSTRACT

In this paper, we have studied Continuous Speech Recognition System of Speaker Adaptation using MAPE (Maximum A Posteriori Probability Estimation) which can adapt any small amount of adaptation speech data. Speaker adaptation is performed by the method of MAPE after Concatenation training which is making sentence unit HMM linked by syllable unit HMM and Viterbi segmentation classifies speech data to be adaptation into segmentation of syllable unit data automatically without hand labelling. For car control speech, the recognition rates of adaptation of HMM was 77.18% which is approximately 6% improvement over that of unadapted HMM.(in case of O(n)DP)

I. 서 론

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 적응화할 수 있는 시스템 개발이 활발히 연구되고 있다. [1,2,3,4] 일반적으로 화자 적응화 방법에는 스펙트럼 특징에 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 데이타 및 적응 인식 시간이 소요되어지며, 후자의 경우는 음성의 변동을 통계적으로 처리하고 이 통계량을 확률형태의 모델에 반영하는 HMM에서 적응화하는 경우이다. HMM에서 Baum-Welch 알고리즘에 의한 최우추정법(Maximum Likelihood Estimation)으로 적응화하는 경우는 표준모델에 적응화 할 학습 데이터를 일괄적으로 주어 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 MAPE법을 이용한 음성 인식 시스템을 개발하였으며, 소규모 문장에서 적용가능성을 검토하였다. 또한 MAPE 추정시 파라메타 적응방법에 따른 인식률을 비교하였다. 본 논문의 구성은 2장에서 최대사후 확률추정법에 의한 화자적응화법을 소개하고, 3장에서는 실시간 음성인식 시스템 구현에 대해 다루었고, 4장에서는 실험 및 결과를 고찰하였고, 5장에서는 결론과 향후 연구과제를 제시하였다.

* 동아대학교 전자공학과
** 섬유기능대학 전자계산기학과
*** 경일대학교 전자정보공학과
접수일자 : 1998년 8월 12일

II. 화자적응화와 최대사후확률 추정법에 의한 파라메터 추정

2.1 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 추정법은 소수의 적응 데이터 만으로 적응화가 가능하며, 순서적 학습에 의해 추가적인 적응화가 가능하다. 본 실험에서는 1개의 학습샘플이 주어질 때마다 최적인 파라메터를 추정하는 최대 사후 확률추정법을 HMM 학습에 이용하였다. 이는 평균 벡터의 추정뿐만 아니라 공분산 행렬의 추정시 1샘플만으로 파라메타 추정이 가능한 순서적 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 구한후 MAPE 추정하는 연결학습법을 이용하여 적응화 실험을 행하였다.[7]

2.2 최대 사후확률 추정법

최대 사후확률 추정법은 Bayesian Successive Estim-ation이라고 부르는 학습법으로서, 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]

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

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

ML의 경우 학습샘플의 조건부 확률이 최대가 되도록 θ를 학습하여 평균벡터와 공분산 행렬을 추정한다. 이 경우 학습시에 모든 샘플을 준비할 필요가 있으며, 학습할 때에는 Baum-Welch 알고리즘을 이용한다. 이 추정방법은 적응화 하기 전의 표준 HMM을 학습할 때에 이용한다.

$$\max_{\Theta} P(X_1,\cdots,X_M|\Theta) \quad (2)$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i \quad (3)$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(X_i-\hat{\mu})(X_i-\hat{\mu})^T \quad (4)$$

평균벡터와 공분산 행렬의 추정법을 다음과 같이 나열하였다.

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

MAPE 추정을 이용하여 다차원 정규분포의 평균벡터를 학습하기 위해서는 식(2)를

$$\theta = \mu \quad (5)$$

로 한다. 여기서 μ는 평균벡터이다.

그리고 1개의 샘플 X1은 $N(\mu,\Sigma)$의 정규분포에 따른다고 가정하고, Σ는 표준 모델의 공분산으로 미리 알고 있는 값으로 하면 식(6)과 같다.

$$P(X_1|\mu) \cong N(\mu,\Sigma) \quad (6)$$

또 사전분포로서, μ는 가장 확실한 평균벡터 $\mu 0$와 불확실성을 나타내는 공분산 행렬 K0의 정규분포에 따른다고 가정하면 식(7)이 된다.

$$P(\mu) \approx N(\mu_0, K_0) \quad (7)$$

이상에서 정의한 확률을 식(2)에 대입하면 식(8)이 된다.

$$\begin{aligned} P(\mu|X_1) &= \frac{P(X_1|\mu)P(\mu)}{\int P(X_1|\mu)P(\mu)d\mu} \approx N(\mu_1, K_1) \\ &= C\exp(-\frac{1}{2}(X_1-\mu)^T\Sigma^{-1}(X_1-\mu) \\ &\quad -\frac{1}{2}(\mu-\mu_0)^T K_0^{-1}(\mu-\mu_0)) \end{aligned} \quad (8)$$

식(8)을 변형하여 μ에 관계하는 항만을 나타낸다. 여기서 C는 μ에 관계없는 항으로 정수이다. 추정된 평균벡터와 불확실성은 식(9)와 같이 된다.

$$\begin{aligned} \hat{\mu}_1 &= K_0(K_0+\Sigma)^{-1}X_1 + \Sigma(K_0+\Sigma)^{-1}\mu_0 \\ \hat{K}_1 &= K_0(K_0+\Sigma)^{-1}\Sigma \end{aligned} \quad (9)$$

K0는 추정전에 μ가 어느 정도의 불확실성을 갖는가를 나타내는 공분산행렬이다. 문헌 [4]에서는 적응화전의 다수화자 모델에서 각 혼합의 평균벡터에서 구한 방법이 소개되어 있지만, 전공분산 행렬을 이용하는 경우는 어느 정도 혼합수가 많은(적어도 10이상 2차원수) HMM이 아니면 정확한 K0를 구하는 것은 불가능하다. 여기에서는 적응화 파라메터 α를 도입하고 K0대신에 실험적으로 구하면 식(10)이 된다.

$$K_0 = \alpha^{-1}\Sigma \quad (10)$$

여기서 α를 0에 근접시키면 K0는 크게 되어 μ의 불확실성이 높고, 역으로 α를 매우 크게하면 K0는 작게되어 μ의 불확실성이 낮다고 가정하는 것이 된다. 식(9)에 대입하면,

$$\hat{\mu}_1 = \frac{\alpha \mu_0 + X_1}{\alpha + 1} \tag{11}$$

이 되고, N 개의 샘플을 반복해서 준 경우의 추정치는 식(12)와 같다.

$$\begin{aligned} \hat{\mu}_N &= \frac{(\alpha + N - 1)\mu_{N-1} + X_N}{\alpha + N} \\ &= \frac{\alpha \mu_0 + \sum_{i=1}^{N} X_i}{\alpha + N} \end{aligned} \tag{12}$$

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

둘째, 평균벡터를 미리 알고 있는 경우의 공분산 행렬 학습은 다음과 같다.

N개의 샘플로 MAPE 추정된 공분산 행렬은 다음과 같이 된다.

$$\begin{aligned} \hat{\Sigma}_N &= \frac{(\alpha + N - 1)\Sigma_{N-1} + X_N X_N^T}{\alpha + N} \\ &= \frac{\alpha \Sigma_0 + \sum_{k=1}^{N} X_k X_k^T}{\alpha + N} \end{aligned} \tag{13}$$

세째, 평균벡터와 공분산 행렬의 동시학습의 경우는 추정해야 할 파라메터가 2개이기 때문에 사전분포와 사후확률은 동시분포가 된다. 1개 및 N개의 샘플에 의해서 추정된 공분산 행렬의 추정치는 다음 식이 되고 평균벡터 추정치는 식(12)와 같다.

$$\begin{aligned} \hat{\Sigma}_1 &= X_1 X_1^T - (\alpha + 1)\mu_1 \mu_1^T + \beta \Sigma_0 + \alpha \mu_0 \mu_0^T \\ \hat{\Sigma}_N &= \frac{1}{\beta + N}\left\{\sum_{i=1}^{N} X_i X_i^T - (\alpha + N)\mu_N \mu_N^T + \Sigma_0 + \alpha \mu_0 \mu_0^T\right\} \\ &= \frac{1}{\beta + N}\{X_N X_N^T - (\alpha + N)\mu_N \mu_N^T + (\beta + N - 1)\Sigma_{N-1} \\ &\quad + (\alpha + N - 1)\mu_{N-1} \mu_{N-1}^T\} \end{aligned} \tag{14}$$

2.3 ML추정 후 파라메터를 MAPE추정.

ML에 의한 순서적 연결 학습법은 적응화용 데이터를 Baum-Welch또는 Viterbi알고리즘으로 추정 학습한후 평균벡터와 분산을 MAPE 추정한 경우이다. 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 벡터를 MAPE 추정으로 선형보간하여 학습한다. 본실험에서는 Viterbi알고리즘에 의해 상태에 대응하는 프레임을 구한후 ML 추정하였다. 하나의 샘플로 HMM 각 상태의 평균벡터를 구할수 있으며, 웨이트 값 α는 10일 때 최적이었다. MAPE를 이용한 화자적응화 과정을 그림 1에 나타내었다

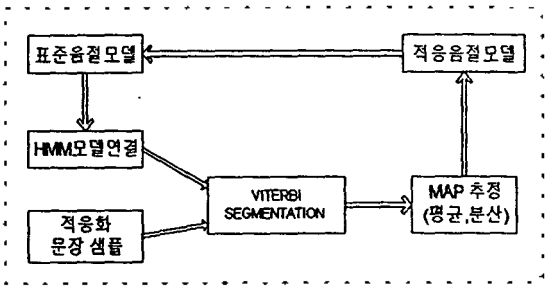

그림 1. MAPE를 이용한 화자적응화 과정
Fig. 1. Process of speaker adaptation using MAPE.

2.4 Viterbi 알고리즘으로 추출한 프레임을 MAPE 추정.

ML 추정된 평균 벡터를 이용하여 MAPE 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 MAPE 추정에 사용하도록 하였다.

III 실시간 연속음성 인식 시스템 구현

3.1 시스템 구성 개요

본 논문에서 구성한 실시간 연속음성 인식 시스템의 블럭도를 그림 2에 보였다.

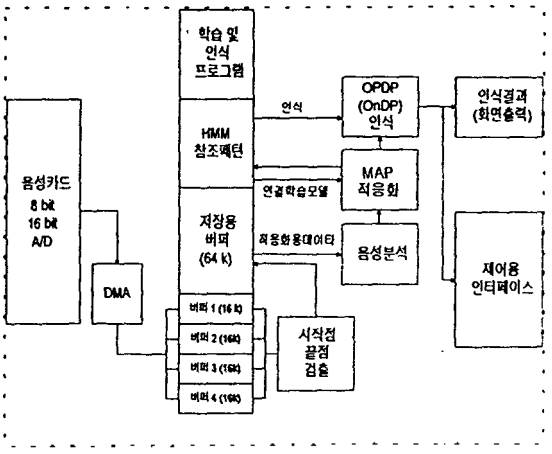

그림 2. 연속음성 인식 시스템도
Fig. 2. Schematic diagram of continuous speech recognition system.

그림 2에서 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.

저장용 버퍼에 저장된 음성을 음성분석 과정에서 10차 멜 켑스트럼을 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 알고리즘으로 패턴 비교하여 인식 결과를 PC 모니터에 문자로 출력함과 동시에 제어용 인터페이스로 제어 데이터를 보낼 수 있도록 구성하였다.

3.2 인식 알고리즘

일반적인 연속음성 인식기술은 기본적으로 인식하고자 하는 음절 또는 단어의 표준 시계열 패턴을 준비하고, 입력된 테스터 패턴과 비 선형적으로 조회하는 DP 매칭법이 오래전부터 연구되고 이용되어 왔다. 특히, 본 실험에 사용된 인식 알고리즘인 O(n)DP와 OPDP법은 입력된 음성 데이터와 학습된 모든 음절 HMM을 DP법으로 매칭을 실시하여 우도가 최대가 되는 최적 음절 또는 단어 계열을 구하는 연속음식 알고리즘이다.

3.2.1 O(n)DP법

Order-n DP(O(n)DP)법은 입력된 음성 데이터의 시계열에 대해 우도가 최대가 되도록 음절 HMM을 연결하고, 연결된 음절 HMM의 모델 심볼을 출력하는 방법이다. 이때 우도의 계산 및 최적인 단어 계열을 구할 때는 viterbi 알고리즘을 이용한다. 그림 3은 O(n)DP법에 의해 인식결과를 구하는 과정으로서 viterbi 알고리즘을 이용하여 최적의 음절계열을 매칭하여, 테스트 패턴의 전 프레임에 대해 참조패턴과 우도값이 최대가 되는 참조패턴의 인덱스를 저장한 후 역 추적을 통하여 인식 결과를 구하는 과정을 나타내었다.

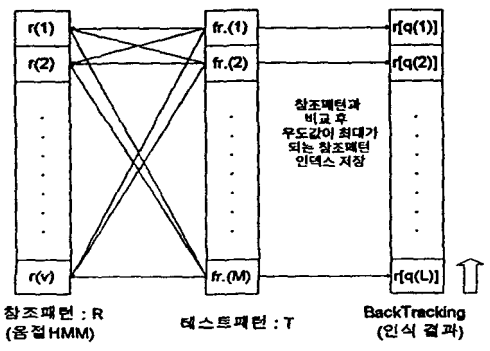

그림 3. DP 매칭 과정
Fig. 3. Process of DP matching.

3.2.2 One Pass DP법

OPDP법은 유한상태 오토마타(Finite State Automata)에 의한 구문의 제어를 통하여 효율적인 탐색을 실시하는 연속음성 인식 알고리즘이다. 구문상의 제약이 많기 때문에 소규모 연속음성 인식에 알맞은 언어모델이다. 구문제어 OPDP법에 의한 연속음성 인식 알고리즘은 O(n)DP법과 기본적으로 같은 DP매칭 알고리즘이며, 오토마타 제어 루프가 추가되어 HMM모델의 모든 단어가 최적 단어(음절)열 검색에 이용되지 않고 현재 진행중인 오토마타 상태 번호까지 구성할 수 있는 부분 문에 등록된 단어(음절)만을 검색하는 방법이다. 그림 4는 본 논문에 사용된 자동차 제어문을 위한 유한상태 오토마타를 나타낸 것이다.

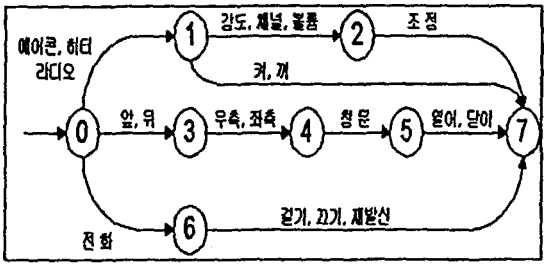

그림 4. 자동차 제어문의 유한 상태 오토마타
Fig. 4. Finite state automata for car control commands.

IV. 인식 실험 및 결과 고찰

시뮬레이션과 실시간으로 구성한 인식시스템에 의한 두가지 방법에 대하여 인식 실험을 행하였다.

시뮬레이션에 의한 본 적응화 실험에서는 기존의 학습된 모델에 다른 화자의 데이터를 추가시켜 화자적응화 실험을 하였다. 적응화 방법으로는 ML로 추정한 파라메터를 가지고 MAPE 추정한 경우와 프레임을 그대로 샘플로 하는 MAPE 추정의 경우에 대해 각각 실험하였다.

실시간 음성인식 시스템에서는 CHMM으로 학습된 모델로 적응화 실험을 행하였고, 인식방법으로 OPDP법을 사용하였으며, 적응화 실험은 프레임을 샘플로 MAPE 추정한 경우에 대해 인식 실험을 행하였다.

4.1 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.

4.2.1 시뮬레이션에 의한 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 5회발성분을 CHMM

모델로 학습하였고, 실험실에서는 녹음한 3사람분의 데이터 중 3회분은 적응화용 데이터로, 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	20ms
특징 파라미터	LPC Cepstrum(14차) -> LPC Melcepstrum(10차)

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

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 MAPE 추정하는 경우에 대해 O(n)DP법으로 인식실험을 하였다. 또한 파라메터 추정방법에 따른 ML 추정후 MAPE 추정하는 경우와 Viterbi 알고리즘으로 추출한 프레임을 MAPE 추정한 경우, 각각에 대해 인식실험을 행하였다. 적응화 실험에서 최적인 적응화 계수를 구하기 위해 적응화용으로 발췌한 문장 중 2문장을 선택하여 α, β값을 임의로 증가시켜 가면서 인식률이 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 치환을 인식 계산에 포함시키지 않았다.

$$\text{인식률} = \frac{\text{인식음절수}}{\text{입력음절수}}$$

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

4.2.2 시뮬레이션 결과

표 3은 적응화 전의 인식률과 적응화 후의 인식률을 비교한 것으로서 상당한 인식률 향상을 얻을 수 있었다. ML 추정후 MAPE 추정하는 실험에서는 평균, 분산 추정시 순서적으로 한 문장이 주어질 때마다 MAPE 추정하도록 하였고, 적응 웨이트값은 음절별로 일률적으로 10을 주었다.

Viterbi로 추출한 프레임을 MAPE 추정한 경우는 문장 전체를 일괄적으로 MAPE 추정하였으며, 웨이트값은 데이터 프레임을 상태별로 세그먼터 한 후 각 상태별로 주었다. 여기서 웨이트값 α, β를 각각 30, 50으로 주어 학습하였다. 실험결과 Viterbi로 추출한 프레임을 평균과 분산을 함께 MAPE 추정한 경우 추정전에 비해 77.18%의 인식률을 얻을 수 있었다.

표 3. O(n)DP 법을 이용한 화자 적응화 인식결과 (CHMM의 경우)
Table 3. recognition rate of speaker adaptation using O(n)DP. (In cae of CHMM).

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

표 4는 10Khz 샘플링, 8bit 양자화한 자동차 제어문을 DDCHMM 모델로 학습한 후 적응화한 경우의 인식률이다.[7] 표 3과 4를 비교해 볼 때 16비트로 양자화한 데이터를 학습한 CHMM모델이 8비트로 양자화한 데이터를 학습한 DDCHMM모델에 비해 인식률이 높았다. 이유는 8비트, 16비트 양자화 표현 수치의 정밀도 및 16비트의 데이터가 8비트 데이타에 비해 비교적 조용한 환경에서 녹음한 것으로 CHMM모델로 학습한 것이 DDCHMM모델에 의해 학습한 것에 비해 인식률이 높은 것으로 판단된다.

4.3 실시간 음성 인식 시스템

시뮬레이션에서 높은 인식률을 얻은 Viterbi로 추출한 프레임을 MAPE 추정하는 방법으로 적응화하고 테스트할 수 있는 실시간 인식 시스템을 개발하여 인식실험 을 행

표 4. O(n)DP 법을 이용한 화자 적응화 인식결과 (DDCHMM의 경우)
Table 4. recognition rate of speaker adaptation using O (n)DP. (In case of DDCHMM).

%	인식	치환	삽입	탈락	seg. rate
적응전 인식결과	54.3	42.7	56.0	3.0	40.4
적응후인식결과 (FRAME-MAPE) 평균+분산추정	71.1	27.2	38.0	1.7	57.2
적응후인식결과 (ML추정후-MAPE) 평균+분산추정	68.4	27.4	41.0	4.2	53.0

하였다. 실시간 실험에서도 마찬가지로 8사람중 5사람은 학습, 나머지 3사람은 적응 및 인식 데이터로서 사용하였다. 적응화 학습 모델로서 DDCHMM을 사용할 경우 CHMM에 비해 학습 및 인식 시간이 길어, 실시간 실험에서의 적응화 학습모델은 CHMM을 사용하였으며, 인식방법은 유한 상태 오토마타 제어인 OPDP법을 이용하여 비교 실험하였다.

표 5. 음성 인식 시스템에서의 화자 적응화 인식결과(OPDP법)
Table 5. recognition rate of speaker adaptation using OPDP in speech recognition system.

%	인식	치환	삽입	탈락	seg. rate
적응전 인식결과	97.50	2.50	0	0	100
적응후인식결과 (FRAME-MAPE) 평균+분산추정	98.50	1.50	0	0	100

표 5에서 적응화후 OPDP법으로 인식한 경우 98.5%로 표 3의 인식결과와 비교하여 상당한 인식률 향상을 얻을 수 있었다.

실시간 화자적응화 실험에서도 프레임을 MAPE 추정한 경우 α, β 웨이트값을 각각 30, 50을 주었다. 그림 4는 자동차 제어문의 유한상태 오토마타를 나타내었다. 여기서 유한상태 오토마타 제어 OPDP법이 O(n)DP법에 비해 삽입 및 탈락의 오인식률이 줄어들고, 구문제어가 비교적 간단한 이유로 인식률이 앞에서 실험한 O(n)DP법보다 향상됨을 알수 있었고, 적응전과 비교해 볼 때 상당한 적응화 효과를 얻을수 있었다.

V. 결 론

본 연구에서는 CHMM을 음절 모델로하여 순서적적으로 주어진 문장을 화자 적응화할 수 있는 알고리즘의 개선과 검토를 연구하였고, 실시간 연속 음성 인식 장치 구현에 대해 연구하였다. 음절 단위를 자동 추출한 후, 1개의 학습 샘플이 주어질 때마다 최대 사후 확률 추정법을 이용하므로 적은 양의 데이터로서도 적응화를 가능하게 하였다.

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

또한 PC에서 실시간 음성인식 시스템을 구축하여 연속음성을 인식실험 하였다. 인식시스템은 마이크로 입력되는 음성을 A/D변환하여, 환상형 버퍼에 순환적으로 저장되도록 구성하였고, 시작점과 끝점을 검출하여 무음구간을 제거한 음성부분만을 저장용 버퍼에 저장하도록 하였다. 저장용 버퍼에 저장된 음성을 음성분석 과정에서 10차 멜 켑스트럼을 구하여 학습 및 인식용 음성 파라메터로 사용할 수 있도록 하였다. 음절단위로 적응화 된 CHMM을 이용한 실시간 음성인식 시스템에서 유한 상태 오토마타를 이용한 OPDP법의 경우 98.5%의 인식률 향상을 얻을 수 있었다.

위의 연구결과 적은 데이터를 가지고 적응화하여 당한 인식률 향상을 얻을 수 있었으며, 보다 인식률 향상을 위해 최적의 적응화 계수를 구하는 효율적인 연구가 이루어져야 할 것이다. 또한 구문제어에 의한 연속음성 인식실험 결과 제한된 소규모의 문장으로 오인식이 발생하더라도 치명적인 영향이 없는 응용분야에서는 연속음성 인식시스템 적용이 가능하리라 생각된다.

참 고 문 헌

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 Learninech of Mixture Densities for Hidden Markov Models," Proc.DARPA Speech and Natural Language Workshop, Pacific Grove, pp 272-277(1991)
6. 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)
7. 김상범, 이영재, 고시영, 허강인, "HMM을 이용한 연속 음성 인식의 화자적응화에 관한 연구", 한국음향학회지, pp5-11, 1995

▲김 상 범(Sang Bern Kim) 1969년 4월 26일생

1993년 2월 : 동아대학교 전자공학과(공학사)
1995년 2월 : 동아대학교 대학원 전자공학과(공학석사)
1998년 2월 : 동아대학교 대학원 전자공학과(박사과정 수료)
1998년 3월~현재 : 섬유기능대학 전자계산기학과 전임강사
※주관심분야 : 음성인식, 합성, 신경망, 멀티미디어

▲김 수 훈(Soo Hoon Kim)

1991년 2월 : 동아대학교 전자공학과(공학사)
1993년 2월 : 동아대학교 대학원 전자공학과(공학석사)
1999년 2월 : 동아대학교 대학원 전자공학과(공학박사)
※주관심분야 : 디지털 신호처리, 음성인식, 신경망, 인공지능

▲허 강 인(Kang In Hur) : 12권 6호 참조