Back-off bigram을 이용한 대용량 연속어의 화자적응에 관한 연구

정회원 최 학 윤*

A Study on Speaker Adaptation of Large Continuous Spoken Language Using back-off bigram

Hak-Yun Choi* *Regular Member*

요 약

본 논문에서는 화자 독립 시스템에서 필요한 화자 적응 방법에 관해 연구하였다. 훈련에 참여하지 않은 새로운 화자에 대해서 bigram과 back-off bigram, MAP와 MLLR의 결과를 비교해 보았다. back-off bigram은 훈련 중 나타나지 않은 bigram 확률을 unigram과 back-off 가중치를 적용하므로 bigram 확률 값에 약간의 가중치를 더하는 효과를 가져온다. 음성의 특징 파라미터로는 12차의 MFCC와 log energy, 1차 미분, 2차 미분을 사용하여 총 39차의 특징 벡터를 사용하였다. 인식 실험을 위해 CHMM, 삼중음소(tri-phones)의 인식 단위, 그리고 bigram과 back-off bigram의 언어 모델을 사용한 시스템을 구성하였다.

key words : bigram; back-off bigram; MAP; MLLR

ABSTRACT

In this paper, we studied the speaker adaptation methods that improve the speaker independent recognition system. For the independent speakers, we compared the results between bigram and back-off bigram, MAP and MLLR. Cause back-off bigram applys unigram and back-off weighted value as bigram probability value, it has the effect adding little weighted value to bigram probability value. We did an experiment using total 39-feature vectors as featuring voice parameter with 12-MFCC, log energy and their delta and delta-delta parameter. For this recognition experiment, We constructed a system made by CHMM and tri-phones recognition unit and bigram and back-off bigrams language model.

I. 서 론

음성은 인간이 사용하고 있는 통신 매체 중 가장 자연스러운 형태이다. 즉, 자신의 의사 표명 혹은 정보의 생성에 있어서 음성을 이용하는 비중이 매우 높다. 음성인식(speech recognition)은 '음성에 포함된 음향학적 정보로부터 음운·언어적 정보를 추출하여 이를 기계가 인지하고 반응하게 만드는 일련의 과정'을 의미한다. 최근에는 낭독체 연속음성인식과 대화체 연속음성인식에 많은 연구를 집중하고 있다. 하지만 음성인식에는 아직 불

* 김포대학 전자정보계열(hychoi@kimpo.ac.kr)
논문번호 : 030109-0317, 접수일자 : 2003년 3월 17일
※ 본 연구는 2002년도 김포대학 특성화(II) '산학공동연구활성화' 프로그램에 의해 연구되었음.

확실한 요소들 즉, 음의 크기 다양한 억양 그리고 화자의 상태 등에 의한 어려움이 많다. 이와 같은 요소들로 인해 음성인식의 적용 범위가 실제의 인간의 능력에 비해 제한 될 수밖에 없다. 이러한 문제를 해결하기 위한 방법으로 화자적응 알고리즘을 사용하여 인식률을 높이려는 연구가 진행 중이다[1,2,3,4].

본 논문에서는 인식 알고리즘으로 CHMM (continuous hidden Markov model)을 사용하여 훈련된 codebook을 만들고, 언어 모델로 back-off bigram을 적용하여 확률 통계적 방법인 MAP(Maximum a posteriori)와 MLLR (Maximum Likelihood Linear Regression)을 이용하여 HMM의 파라미터 적응시키는 방법으로 대용량 화자 적응 알고리즘을 구현하였다.

2장에서는 음성인식 시스템, 언어모델, 본 논문에서 사용한 back-off bigram을, 3장에서는 적응알고리즘을, 4장에서는 인식실험 및 결과 그리고, 5장에서는 결과 분석 및 향후 연구 과제로 본 논문을 구성하였다.

II. 연속 음성 인식 시스템

2.1 음성자료 수집

표준어를 구사하는 20대 초반에서 30대 초반의 남성과 여성 화자 82명에 대해서 음성 자료를 구성하였다. 신문 낭독 체를 서로 다른 5문장씩 발음하여 총 410문장의 음성 데이터가 사용되어 졌다. 모든 음성 데이터는 16 kHz의 sampling rate와 16 bits의 해상도로 녹음하여 Database로 사용하였다. 특징벡터로는 12차의 MFCC와 1차의 Energy, 그리고 이들을 1차 미분한 delta계수 및 2차 미분한 delta-delta계수등 총 39차의 계수를 사용하였다.

2.2 분석방법

마이크를 통해 발성된 음성신호를 A/D 변환기를 통해 이산 데이터로 변환하였다.

Pre-emphasis 필터와, 32ms Hamming window를 사용하였고, Overlap은 16ms로 하였다. 사용된 특징 파라미터는 12차 MFCC, 1차 Energy와 각각의 delta-Cepstrum 및 delta-Energy, delta-delta-Cepstrum, delta-delta-Energy 계수를 사용하였다[15].

표 1. 파라미터 값

Table 1. Parameter Value.

Parameter	Value
Pre-emphasis constant	0.95
Window size	Hamming Windowing 512 sample
Window overlap size	256 sample
Coefficients	12-th order MFCC + 1-th order energy + D + DD = 39-th order delta delta MFCC
Probability density function	Diagonal Gaussian mixture
Topology of model Q	Left to right model

2.3 연속 음성 인식 시스템

음성이 사람과 컴퓨터 사이에서 자연스러운 인터페이스로 자리 잡으려면 무엇보다도 자유도, 즉 대용량의 어휘와 그것을 이용한 다양한 문장표현을 처리해야 한다. 대부분의 연속음성인식시스템은 통계적 모델에 기반을 두고 있고, 음향 모델의 초기화, 음향 모델들의 훈련, 언어모델의 훈련 이 세 가지 과정은 필수적이다.

다음은 본 실험에서 구현된 연속음성인식 시스템의 구성도 이다.

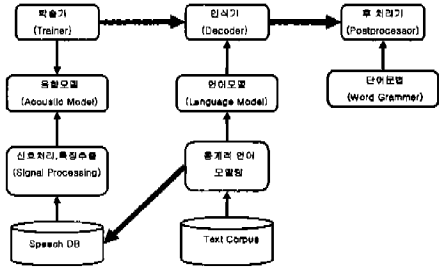

그림 1. 연속 음성 인식 시스템 구성도

Fig. 1. Continuous speech recognition system.

1) 음성신호처리

본 논문에서는 특징벡터로 12차의 MFCC와 1차의 Energy, 그리고 이들을 1차 미분한 delta계수 및 2차 미분한 delta-delta계수등 총 39차의 계수를 사용하였다[15].

2) Mel Frequency Cepstrum Computation

MFCC는 filer bank energy에 log를 취한 후

다시 IDFT를 취하여 구할 수 있다. filter bank에서 얻어진 출력 $Y_t(m)$에 logarithm을 취하고 inverse DFT를 수행하여 특징벡터 MFCC $y_t^{(m)}(k)$을 구한다.

$$y_t^{(m)}(k) = \sum_{m=1}^{M} \log(|Y_t|)\cos\left(k\left(m-\frac{1}{2}\right)\frac{\pi}{M}\right), \quad k=0,\ldots,L \tag{1}$$

여기서 M은 filter bank 채널들의 수를 말하고, 0차 MFCC계수 $y_t^{(0)}(k)$는 해당 프레임에서의 log 에너지와 유사하다[15].

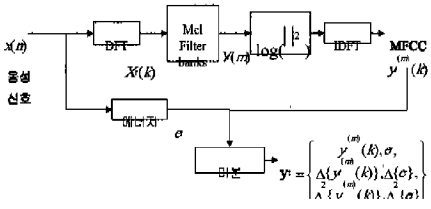

그림 2. MFCC 과정도
Fig. 1. MFCC Processing.

3) Delta Coefficients and Energy Measures

일반적으로 켑스트럼 계수의 경우 프레임별 logarithm energy와 함께 사용되어지는데 앞에서 설명한 것처럼 인식기의 성능 향상을 위해 1차 및 2차 미분을 사용한다[8]. 본 논문에서는 12차의 MFCC와 1차의 Energy, 그리고 이들을 1차 미분한 delta계수 및 2차 미분한 delta-delta계수등 총 39차의 계수를 사용하였다[15].

$$\Delta^i\{\mathbf{u}_t\} = \Delta^{i-1}\{\mathbf{u}_{t+1}\} - \Delta^{i-1}\{\mathbf{u}_{t-1}\}$$
$$\Delta^0\{\mathbf{u}_t\} = \mathbf{u}_t \tag{2}$$

2.4 인식 알고리즘

HMM에서의 관측 symbol은 음성신호의 특징을 나타내는 특징벡터가 되는데, DHMM의 경우 관측 symbol이 VQ에 의해서 양자화된 특징벡터들을 사용하고, CHMM(Continuous Hidden Markov Model)의 경우 특징벡터들이 연속적인 분포를 가지게 함으로써 그들의 평균과 분산을 이용하여 관측한 symbol이 연속 분포함수를 가지게 한다. 이렇게 연속분포함수를 가지는 관측 symbol은 이산값에 비해 매우 정확한 상태표현을 가능하게 한다. CHMM에서는 $b_j(\mathbf{y})$을 확률밀도함수로 정의하여 사용한다.

확률 밀도 함수는 다음의 식으로 주어진다.

$$b_j(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^D delU_t}} e^{\left\{-\frac{1}{2}(\mathbf{y}-\mu_j)^T U_j(\mathbf{y}-\mu_j)\right\}} \tag{3}$$
$$j=1,\ldots,N$$

$\mathbf{y}_t$: n 차원의 정규 분포를 가지는 출력벡터
μ : D 차원의 평균벡터
U_j : covariance matrix

CHMM에서 full covariance matrix를 사용하려하면 많은 계산 량이 요구된다. 이는 보통 켑스트럼(Cepstrum)계수의 uncorrelation 성질을 사용하여 계산 량을 효과적으로 줄일 수 있도록 하는 diagonal covariance matrix를 사용하여 해결한다. Diagonal covariance matrix를 U_j라 하면 확률밀도함수는 다음 식과 같이 쓸 수 있다.

$$b_j(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^D \prod_{i=1}^{D}\sigma_{ji}^2}} e^{\left\{-\frac{1}{2}\sum_{i=1}^{D}\frac{(y_i-\mu_{ji})^2}{\sigma_{ji}^2}\right\}}$$
$$= \frac{1}{G_j} e^{-\xi_j(\mathbf{y})} \tag{4}$$

y_i : 특징벡터
μ_{ji} : 상태 $s_t=j$ 일 때 i번째 평균벡터
σ_{ji}^2 : 상태 $s_t=j$ 일 때 i번째의 공분산

$$G_j = \sqrt{(2\pi)^D \prod_{i=1}^{D}\sigma_{ji}^2}$$

$$\xi_j(\mathbf{y}) = \frac{1}{2}\sum_{i=1}^{D}\frac{(y_i-\mu_{ji})^2}{\sigma_{ji}^2}$$

다양한 화자의 음성을 표현하려면 위의 단일 모델을 변형 없이 그대로 사용할 수 없다. 이 문제는 pdf의 mixture를 사용함으로써 해결할 수 있다. 즉 단일 모델에 가중치 w_j^n을 곱한 pdf들의 선형결합을 이용함으로써 다양한 화자의 음성표현이 가능케 된다. 여기서 w_j^n는 j번째 n개의 가중치이다.

$$b_j(\mathbf{y}) = \sum_{n=1}^{M} w_j^n b_j^n(\mathbf{y}) \quad \sum_{n=1}^{M} w_j^n = 1 \quad w_j^n \le 1 \qquad (5)$$

위의 식에서 M은 mixture 개수이다[15].

2.5 언어 모델

연속 음성 인식 시스템에서 이웃하는 단어 사이의 연관성, 즉 언어 정보는 매우 중요한 역할을 한다. 따라서 연속음성 인식에서 언어 모델은 필수적으로 구현이 되어야 하며, 그 인식 영역에 따라서 적절한 언어 모델을 선정하여야 한다[10][12]. 식(6)에서 P(Y|W)는 음향 모델이고 P(W)는 언어 모델이다. 언어 모델은 구문론(syntax)과 어의론(semantics)을 적용할 수 있다. 이때 구문론만을 적용할 경우 보통 언어 모델을 문법(grammar)이라고 부르며 언어 모델을 제안된 영역에서 확정 지어 놓은 것을 Finite State Network(FSN)이라고 한다.

$$\hat{W} = ArgMax_{\,w} P(Y|W)P(W) \qquad (6)$$

여기서 $\hat{W}$은 인식된 단어 열이고, W는 발음된 단어 열이다.

언어 모델의 목적은 임의의 단어 열 W가 주어진 영역에서 어느 정도의 확률을 갖는지 알아내는 것이다. W를 특정한 단어 열이라고 가정했을 때, 확률 P(W)의 계산은 식(7)과 같다.

$$\begin{aligned} P(W) &= P(w_1, w_2, \ldots, w_L) \\ &= P(w_1) \prod_{i=2}^{L} P(w_i|w_1, \ldots, w_{i-1}) \end{aligned} \qquad (7)$$

그런데 이러한 계산은 실제로 거의 불가능하다. 그래서 통상적으로 bigram 이나 trigram 등의 N-gram 단어 모델을 사용한다. N-gram 단어 모델은 식(8)으로 나타낸다.

$$\begin{aligned} P_M(W) &= P(w_1, w_2, \ldots, w_{N-1}) \\ &\times \prod_{i=N}^{L} P(w_i|w_{i-N+1}, \ldots, w_{i-1}) \end{aligned} \qquad (8)$$

즉, N-gram 단어 모델은 I번째 단어의 발생확률을 계산하는데 이전에 발생된 N-1개의 단어만을 이용하여 계산하는 것이다. 그러나 학습 데이터로부터 모든 단어 쌍에 대한 확률 값을 계산할 수는 없으므로 back-off 된 확률 값을 사용한다. 가령 bigram 확률인 $P(w_i|w_j)$가 나타나지 않았다면, 대신 unigram인 $P(w_i)$를 사용한다. 그리고 이 확률 값에 w_j가 w_i의 bigram 관계자가 아님을 뜻하기 위해 back-off 가중치를 적용한다.

2.6 Back-off bigram

본 연구에서는 언어 모델로 back-off bigram을 사용하였다. Back-off bigram을 이용하면 학습 시에 나타나지 않은 단어 쌍이 인식과정 가운데 나타나는 경우, bigram 대신 unigram과 back-off 가중치를 결합한 값을 대신 사용한다.

N(i,j)을 단어 i, j가 bigram 관계를 나타난 회수로 정의하고, N(i)을 학습 문장 가운데 단어 i가 나타난 회수로 정의했을 때, unigram 확률 p(i)는 다음과 같이 정의된다.

$$p(i) = N(i)/N \qquad (9)$$

그리고 back-off bigram은 다음과 같이 구할 수 있다.

$$p(i) = \begin{cases} (N(i,j) - D)/N(i) & , N(i,j) > 0 \\ b(i)p(i) & , otherwise \end{cases} \qquad (10)$$

여기서 b(i)가 back-off 가중치로, 모든 p(i, j)의 합이 1이 되도록 보장하며, 다음 식과 같이 계산된다.

$$b(i) = \frac{1 - \sum_{j \in B} p(i,j)}{1 - \sum_{j \in B} p(j)}, \quad (B = i/all \quad j, p(i,j)) \qquad (11)$$

그리고 D는 discount 상수로, 이 상수를 사용하면 bigram 확률 값의 일부를 이보다 적게 나타나는 bigram 확률 값에 약간의 가중치를 더하는 효과를 가져온다[13].

III. 화자 적응 시스템

일반적으로 음성인식에 있어서 적응기법은 화자 적응이 주가 되고 있다. 즉, 화자 독립 시스템에 특정화자에 대한 음성 데이터를 적응시켜 특정 화자에 대한 인식 성능을 높이기 위한 것이 목적이다. 이러

한 화자의 모델을 적응하는 기법에는 DMA(Direct Model Adaptation)과 IMA(Indirect Model Adaption)이 있다. DMA에서 대표적인 방법으로 Maximum A Posteriori(MAP)가 있고, IMA에서 대표적인 방법으로는 Maximum Likelihood Linear Regression(MLLR)이 있다.

3.1 MLLR(Maximum Likelihood Linear Regression)

IMA 기술은 모델 파라미터의 클러스터링에 관한 일반적인 변환(transformation)을 이용하는 방법이다. 이는 각각의 모델들이 모두 변환되기 때문에 적은 양의 적응 데이터를 사용하고자 할 경우에 매우 효과적이다. 일반적으로 IMA는 식(12)과 같이 클러스터 c에 속하는 모든 모델 파라미터 λ_c가 동시에 함수 $F_\eta(\cdot)$에 의해 변환되어지기 때문에 전역(global) 적응기술 또는 transformation-based adaptation 이라고 한다[16].

$$\lambda'_c = F_\eta(\lambda_c) \tag{12}$$

변환 파라미터 η는 일반적으로 ML(Maximum Likelihood)를 거쳐서 식(13)과 같이 평가된다.

$$\eta_{ML} = ArgMax_\eta P(Y|\lambda, \eta) \tag{13}$$

여기서 $P(Y|\lambda, \eta)$는 변환된 모델 $F_\eta(\lambda)$을 사용하여 얻은 적응 데이터의 Likelihood를 나타낸다.

$$P(Y|\lambda, \eta) = P(Y|F_\eta(\lambda)) \tag{14}$$

변환이 HMM mean vector들의 유사변환(affine transformation)일 때, 이러한 접근을 MLLR이라고 한다.

3.2 MAP(Maximum a posteriori)

DMA 기술은 IMA기술처럼 변환 함수에 의해서 모델 파라미터를 수정하지 않고, 적응 데이터를 사용하여 직접적으로 추정한다.

이러한 DMA 기술 중에서도 Baysian 파라미터 훈련이 대표적인 접근방법인데, 관측되지 않는 모델 파라미터들은 수정되지 않고 단지 적응 데이터 안에서 관측된 모델 파라미터만 수정되기 때문에 Baysian learning은 local 적응기술이라고 한다. Baysian learning은 MAP 평가에 의해 수행될 수 있다[16].

$$\lambda_{MAP} = ArgMax_\lambda P(\lambda|y) \tag{15}$$

여기서 $P(\lambda|y)$는 관측열 Y가 주어진 모델 파라미터 λ의 사후 확률밀도함수(a posterior pdf)이다. MAP에서는 모델 파라미터 λ가 사전분포(prior distribution)이라고 불리는 확률밀도함수 $G(\lambda)$에 의해 기술된 임의의 벡터라고 가정한다. Bayes rule을 사용하여 식(15)을 식(16)과 같이 나타낼 수 있다.

$$\lambda_{MAP} = ArgMax_\lambda \frac{P(\lambda|y)G(\lambda)}{P(Y)} = ArgMax_\lambda P(\lambda|y)G(\lambda) \tag{16}$$

다음은 본 실험에서 구현된 화자적응 시스템의 구성도 이다.

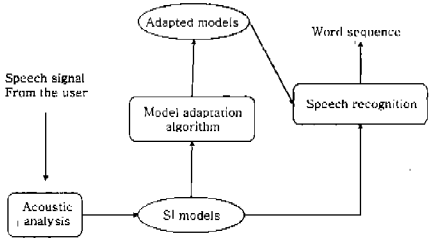

그림 3. 화자적응 시스템의 구성도
Fig. 3. Block diagram of speech adaptation system .

Ⅳ. 인식 실험 및 결과

4.1 음성 데이터베이스

대용량 음성인식기를 구현하기 위해 일반 일간지 기사를 낭독하여 자료를 수집하고 인식 실험을 수행하였다. 학습 데이터로써 표준어를 구사하는 20대 초반에서 30대 초반의 남성과 여성 화자 82명에 대해서 구성하였다. 신문 낭독체를 서로 다른 5문장씩 발음하여 총 410문장의 음성 데이터가 사용되어 졌다. 모든 음성 데

이터는 16 kHz의 sampling rate와 16 bits의 해상도로 녹음하여 Database로 사용하였다. 특징벡터로는 12차의 MFCC와 1차의 Energy, 그리고 이들을 1차 미분한 delta계수 및 2차 미분한 delta-delta계수등 총 39차의 계수를 사용하였다.

4.2 인식 단위 학습

한국어의 음운 현상을 고려하여 묵음 및 잡음을 포함한 총 47개의 단일 음소를 선정하고 그 단일음소로부터 삼중음소의 목록을 만들었다. 먼저 단일음소를 훈련시킨 후 총 3267개의 삼중음소 모델을 만들었다.

4.3 실험 결과

Gaussian mixture를 2개에서 10개까지 늘려 실험하였다. 그 결과 10개의 mixture를 사용한 음향모델이 우수한 인식결과를 나타내었고, 10개의 Gaussian mixture를 갖는 음향모델을 구성하여 bigram과 back-off bigram으로 나누어 단어 인식실험을 수행하였다. 인식결과는 Percent Correct와 Percent Accuracy로 나타내었다.

식(17)은 Percent Correct, 식(18)은 Percent Accuracy를 식(19)은 Word Error Rate를 나타낸다.

$$Percent\ Correct(\%) = \frac{N-D-S}{N} \times 100\% \quad (17)$$

$$Percent\ Accuracy(\%) = \frac{N-D-S-I}{N} \times 100\% \quad (18)$$

$$WER(\%) = \frac{D+S+I}{N} \times 100\% \quad (19)$$

여기서 Accuracy는 단어인식률, N_{sol}은 실제의 단어 수, I는 삽입된 단어 수, O는 탈락된 단어 수, S는 바뀐 단어 수를 나타낸다.

〔표 2〕는 화자적응을 하기 전의 훈련에 참여한 화자(SD model)에 대한 mixture와 관계된 인식률을 나타내며 〔표 3〕은 화자적응을 한 후의 새로운 화자(SI model)에 대한 bigram 인식결과를 보여준다. 또한 〔표 4〕는 화자적응을 한 후의 새로운 화자(SI model)에 대한 back-off bigram 인식결과를 보여준다.

표 2. 훈련된 데이터의 mixture개수에 따른 인식결과
Table 2. Recognition results of mixture coefficient for practice data.t

mixture 개수	Corr.	Acc.	WER
2개	80.88	75.61	24.39
4개	86.60	83.31	16.69
6개	91.15	89.03	10.97
8개	93.89	92.45	7.55
10개	95.67	95.83	4.37

표 3. 적응한 후의 bigram 단어 인식률
Table 3. Word recognition rate of bigram after adaptation processing.

Acc(%)	MAP	MAP 평균	MLLR	MLLR 평균
화자 A	57.35	58.69	64.74	63.68
화자 B	61.12		63.36	
화자 C	58.41		62.16	
화자 D	56.72		65.24	
화자 E	59.84		62.87	

표 4. 적응한 후의 back-off bigram 단어인식률
Table 4. Word recognition rate of back-off bigram after adaptation processing.

Acc(%)	MAP	MAP 평균	MLLR	MLLR 평균
화자 A	63.15	63.43	68.99	68.14
화자 B	64.32		65.86	
화자 C	63.81		67.35	
화자 D	62.25		69.71	
화자 E	63.64		68.8	

V. 결과 분석 및 향후 연구 과제

〔표 2〕를 보면 mixture의 개수가 많아질수록 단어 인식률이 높아짐을 볼 수 있다. 그러나 mixture의 개수가 늘어날수록 많은 계산 량이 필요하다. 따라서 음향모델을 선정할 때 계산 량과 인식률의 관계를 고려하여 선정할 필요가 있다. 〔표 3〕과 〔표 4〕의 결과는 각각 화자적응을 수행한 후의 새로운 화자에 대한 bigram과 back-off bigram의 단어 인식결과를 나타낸다. MAP의 경우 bigram과 back-off bigram의 인식 결과를 볼 때 back-off bigram을 사용했을 때의 단어 인식률이 4.74% 높아졌음을 볼 수 있고, MLLR의 경우 4.46% 인식률이 향상된 것을 볼 수 있다. 즉 평균 4.6%정도 단어인식률이 향상된 것을 볼 수 있다. back-off

bigram은 훈련 중 나타나지 않은 bigram 확률을 unigram과 back-off 가중치를 적용하므로 bigram 확률 값에 약간의 가중치를 더하는 효과를 가져온다. 이와 같은 작용은 back-off bigram의 인식결과가 bigram 인식 결과 보다 향상될 수 있음을 나타낸다[13].

MLLR의 경우 각각의 모델들이 모두 변환되기 때문에 적은 양의 적응 데이터를 사용하고자 할 경우에 매우 효과적이다. MAP의 경우 관측되지 않는 모델 파라미터들은 수정되지 않고 적응 데이터 안의 관측된 모델 파라미터만 수정되기 때문에 local 적응기술이라 한다. MLLR과 MAP의 장점을 혼합한 화자적응 모델을 구성되어질 수도 있다. 또한 화자 분류 방식(Speaker Clustering Approaches)을 이용한 방법도 모색되고 있는데, 대표적인 화자 분류 방식은 남성과 여성을 분류한 후 적응을 수행하는 방법이 대표적인 화자 분류 방식중 하나이다 [14].

참 고 문 헌

[1] W. A. Rozzi and R. M. Stern, "Speaker adaptation in continuous speech recognition via estimation of correlated mean vectors", in Proc, ICASSP, pp.865-868, 1991.

[2] X. D Huang and K. F. Lee. "On speaker independent, speaker dependent, and speaker adaptive speech recognition", in Proc. ICASSP, pp. 877-880, 1991.

[3] Y. Zhao, "An acoustic-phonetic-based speaker adaptation technique for improving speaker independent continuous speech recognition", IEEE Trans. Speech Audio Processing, vol.2. no.3. pp 380-394, July. 1994.

[4] B. F. Necioglu, M. Ostendorf, and J. R. Rohlicek, "A bayesian approach to speaker adaptation for the stochastic segment model", in Proc, ICASSP, pp.437-440, 1992.

[5] L. R. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentic Hall, 1993.

[6] Oppenheim A. V. and Shafer R.W.,*Digital Signal Processing*, Prentice Hall
. 1989.

[7] Jayant N.O.S., Noll P., *Digital Coding of Waveforms*, Prentice Hall, 1984.

[8] Fruri S., "Cepstral analysis techniques for automatic speaker verification", IEEE Tran. on ASSP, 29, No.2, pp.254-272, 1981.

[9] Rabiner L. R., Shafer R.W., *Digital Processing of Speech Signal*, Prentice Hall, 1993.

[10] Ney H., Essen U., Kneser R., *"On structuring probabilistic dependencies in stochastic language modeling", Computer Speech and Language*, vol. 8, pp.1-38, 1994.

[11] Niles L.T., "Modelling and learning in speech recognition : the relationship between stochastic pattern classigiera and neural networks", Technical Report LWMS-79, Brown Unibersity, Aug. 1990.

[12] Ney H., "Language models", in Handbook of Standards and Resources for Spoken Language Systems, Mouton de Gruyter, 1997.

[13] Steve Young, and et al, The HTK Book. 2000

[14] Xueding Huang., Alejandre Acero., Hsiao-Wuen Hon., "Spoken Language Processing", Carnegie Mellon University, pp.446-451

[15] Claudio Becchetti., Lucio Prina Ricotti., "Speech Recognition", John Wiley & Sons Ltd. pp.212-228, 1999.

[16] W. Chou, "Maximum a posterior linear regression with elliptically symmetric matrix variate priors." in *Proc. Eur. Conf. Speech Communication Technology*, Budapest, Hungary, 1999.

최 학 윤(Hak-Yun Choi) 정회원

1985년 2월 : 숭실대학교 전자공학과 졸업
1987년 2월 : 숭실대학교 전자공학과 석사
1999년 8월 : 건국대학교 전자공학과 공학박사
1996년 3월~현재 : 김포대학 전자정보제열 조교수

〈주관심분야〉 마이크로파, 신호처리