

# 확률적 매칭 방법을 사용한 음소열 기반 음성 인식

## Phonetic Transcription based Speech Recognition using Stochastic Matching Method

김 원 구

Weon-Goo Kim

군산대학교 전자정보공학부

생체인식연구센터

### 요 약

본 논문에서는 화자 독립 음소 인식기를 사용하는 음소열 기반 음성 인식 시스템의 성능을 향상시키는 새로운 방법을 제안하였다. 화자독립 음소 HMM을 사용하는 음성 인식 시스템은 입력 문장에 대한 음소열만을 사용하므로 저장 공간은 크게 줄일 수 있다. 그러나 시스템의 성능은 화자독립 모델을 사용하므로 발생하는 음소 오차 때문에 화자 종속 시스템보다 저하된다. 여기에서는 화자 적응 기술을 사용하여 화자독립 모델과 학습 데이터 간의 불일치를 감소시키도록 음소열과 변환 벡터를 반복적으로 추정하는 학습 방법을 제안하였다. 화자 적응을 위한 변환 벡터를 추정하기 위하여 확률적 매칭 방법이 사용되었다. 실험은 전화선을 통하여 얻어진 데이터를 사용한 실험에서 기존 방법에 비하여 약 45%정도 오차가 감소되었다.

키워드 : 음성 인식, 확률적 매칭, 음소 HMM, 음소열

### Abstract

A new method that improves the performance of the phonetic transcription based speech recognition system is presented with the speaker-independent phonetic recognizer. Since SI phoneme HMM based speech recognition system uses only the phoneme transcription of the input sentence, the storage space could be reduced greatly. However, the performance of the system is worse than that of the speaker dependent system due to the phoneme recognition errors generated from using SI models. A new training method that iteratively estimates the phonetic transcription and transformation vectors is presented to reduce the mismatch between the training utterances and a set of SI models using speaker adaptation techniques. For speaker adaptation the stochastic matching methods are used to estimate the transformation vectors. The experiments performed over actual telephone line shows that a reduction of about 45% in the error rates could be achieved as compared to the conventional method

Key Words : speech recognition, stochastic matching, phonetic HMM, phonetic transcription

### 1. 서 론

일반적으로 음성 다이얼링 시스템은 화자 종속형의 시스템을 사용하여 각 화자가 자동적으로 전화를 걸 때 사용할 명령어나 키워드를 포함하는 개인적인 목록을 사용한다. 이러한 시스템의 구성은 보통 다음과 같은 두 가지 과정을 거친다. 먼저, 등록 단계에서 사용자는 각각의 단어나 문장을 수 차례 발성하고 그에 해당되는 전화번호를 제공한다. 생성된 모델은 인식을 위하여 저장된다. 다음으로 인식 단계에서는 사용자의 신원이 확인된 후 사용자는 단어 또는 문장을 발음하고 시스템은 인식을 수행하여 자동으로 해당된 전화번호로 전화를 건다. 이러한 형태의 시스템은 그 구조가 간단

하고 화자종속의 형태를 갖기 때문에 인식 성능이 비교적 우수하지만 단어나 문장 단위로 모델을 저장해야 하기 때문에 저장 공간이 많이 필요하고 인식 대상 단어수의 증가에 비례하여 필요한 저장도 증가하게 된다. 이러한 문제점은 핸드폰에 사용되는 음성 다이얼링 시스템과 같이 한 명의 사용자가 수십 단어 정도를 사용하는 경우에는 큰 문제가 되지 않지만 전화망이나 네트워크를 사용한 음성 다이얼링인 경우와 같이 수십 또는 수백만 명의 데이터를 서비스 사업자의 서버에 저장해야 하는 경우에는 음성인식을 수행하기 위한 데이터 저장 공간의 크기가 매우 커지기 때문에 중요한 문제가 된다.

이러한 문제를 해결하기 위한 방법 중의 하나로 화자독립 음소모델을 이용한 방법들이 제안되었다[1-4]. 이러한 방법들은 화자독립 음소모델을 사용하여 학습 데이터의 음소열을 구한 후 음소열을 저장하고, 입력 음성을 인식할 때 저장된 음소열과 화자독립 모델을 사용하는 것이다. 이러한 방법들의 장점은 저장 공간은 크게 줄일 수 있으나 다음과 같은 두 가지 문제점을 가지고 있다. 첫 번째는 화자독립 음소 HMM을 사용한 음소열 추정결과에 많은 오차가 발생하는

접수일자 : 2007년 4월 18일

완료일자 : 2007년 7월 31일

감사의 글 : 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (R11-2002-105-04004-0(2007))

것이다. 두 번째는 화자독립 모델을 음소 인식에 사용할 때 발생하는 오차로 인하여 화자종속 모델을 사용하는 방법보다는 인식 성능이 저하되는 문제점이 있다.

본 논문에서는 화자독립 음소 모델을 사용한 음성 다이얼링 시스템의 성능을 개선하기 위하여 음소 열과 화자적응을 위한 모델 변환 함수를 동시에 추정하는 방법을 제안하였다. 여기서 화자적응을 위한 변환 벡터는 확률적 매칭(stochastic matching) 방법을 위한 최고 유사도(maximum likelihood) 방법[5,6]을 이용하였으며 음소 열과 함께 반복적으로 추정되었다. 이러한 변환 벡터는 크기가 작아서 작은 저장 공간을 사용하면서도 인식 성능을 화자종속 시스템에 근사하도록 향상시킬 수 있었다.

## 2. 확률적 매칭을 사용한 학습 알고리즘

본 논문에서는 화자독립 음소 모델을 사용한 음성 다이얼링 시스템의 성능을 향상시키기 위하여 화자적응 방법을 사용하여 성능을 향상시키는 방법을 제안하였다. 기존 화자독립 모델을 사용하는 음성 다이얼링 시스템은 학습단계에서 등록에 사용되는 음성으로부터 음소인식을 수행하여 음소 열을 구한 후 이 음소 열을 저장한다[1-4] 인식 단계에서는 저장된 음소 열과 화자독립 음소 HMM을 연결한 모델을 만든 후 입력 음성에 대한 확률을 구한다. 이러한 방법은 저장해야 할 데이터가 음소 열이므로 필요한 저장공간이 매우 작아지는 장점이 있다. 그러나 이러한 방법은 저장공간은 크게 줄일 수 있으나 화자독립 모델을 음소 인식에 사용할 때 발생하는 오차로 인하여 화자종속 모델을 사용하는 방법보다는 인식 성능이 저하되는 문제점이 있다.

본 논문에서는 화자독립 음소 모델을 사용한 음성 다이얼링 시스템의 성능을 개선하기 위하여 음소 열과 화자적응을 위한 모델 변환 함수를 동시에 추정하는 방법을 제안하였다. 제안된 시스템의 구조는 그림 1과 같다.

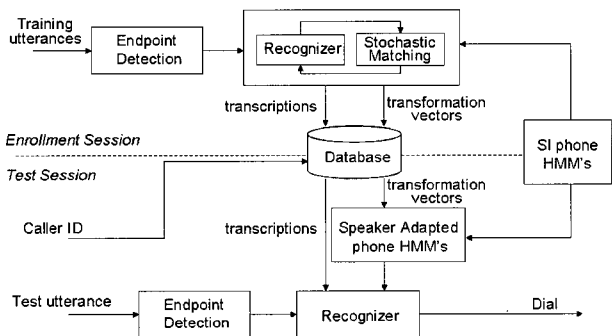


그림 1. 제안된 시스템의 개념도.  
Fig. 1. The concept of the proposed system.

제안된 방법은 등록 단계인 학습과정에서 학습 데이터와 화자독립 음소 HMM을 사용하여 학습 데이터의 음소 열과 화자적응을 위한 변환 벡터(bias)를 동시에 추정한 후 음소 열과 함께 저장하고, 인식 단계에서 화자독립 음소 HMM을 각 화자의 변환벡터를 사용하여 변환한 후 입력 음성에 대한 인식을 수행한다. 여기서 화자적응을 위한 변환 벡터는 확률적 매칭을 위한 최고 유사도 방법을 이용하였으며 음소 열과 함께 반복적으로 추정되었다[5,6].

확률적 매칭을 위한 최고 유사도 방법은 다음과 같이 적

용될 수 있다. 우선 일련의 특징 벡터  $Y = \{y_1, y_2, \dots, y_T\}$ 와 화자독립 음소 HMM의 집합을  $\Lambda_X$  라고 할 때, 화자 적응되어 변형된 모델  $\Lambda_Y$ 를 위한 모델 공간 변환은  $\Lambda_Y = G_\eta(\Lambda_X)$ 로 이루어진다. 여기서  $G_\eta(\cdot)$ 는 모델 변환 함수이고  $\eta$ 는 변환 파라미터이다.  $Y$ 와  $\Lambda_X$  사이의 불일치를 줄이는 방법중의 하나로 변환 파라미터  $\eta$ 와 음소열  $W$ 의 결합 유사도를 최대화하는  $\eta$ 와  $W$ 를 찾는 것이고 다음과 같이 정의된다.

$$\begin{aligned}
 (\eta', W') &= \operatorname{argmax}_p(Y, W | \eta, \Lambda_X) & (1) \\
 &(\eta, W) \\
 &= \operatorname{argmax}_p(Y | W, \eta, \Lambda_X)P(W) \\
 &(\eta, W)
 \end{aligned}$$

식 (1)에서 변수  $\eta$ 와  $W$ 에 대한 결합 최대화는 먼저  $\eta$ 를 고정시키고  $W$ 에 대하여 최대화하고, 그 다음에  $W$ 를 고정시키고  $\eta$ 에 대하여 최대화하는 반복적인 방법으로 구현될 수 있다. 본 논문에서는 음소 군(phonetic class)과 코드북(codebook) 기반의 두 가지 종류의 확률적 매칭 방법을 사용하여 변환 벡터를 추정하였다. 음소 군 기반 확률적 매칭 방법에서는 음소 군에 따라 다수의 변환 벡터가 사용되었다. 변환 벡터의 개수는 음향학적 음소 군에 따라 결정되었다. 코드북 기반의 확률적 매칭 방법에서는 모델 파라미터를 공통으로 사용하는 묶음(tying)의 개념을 사용하여 변환벡터의 개수를 결정하였다. 묶음의 정도에 따라 코드북의 크기와 그에 따른 변환벡터의 수가 조절되어진다. 코드북의 크기는 화자독립 전체 화자독립 HMM에 공통으로 적용되는 하나의 변환벡터를 사용하는 한 개에서부터 가능한 혼합 성분의 수까지 가능하다. 코드북은 유클리디안(Euclidean) 거리측정 방법을 사용하는 Lloyd 알고리즘을 사용하여 화자독립 HMM의 모든 혼합 성분들의 평균 벡터를 군집화하여 구성되었다.

$\Omega_1, \dots, \Omega_K$ 를  $K$ 개의 군집으로 가정하면, 음소군 또는 코드북 기반 확률적 매칭에서는  $K$ 는 음소 군 또는 코드 워드의 개수가 된다. 변환 벡터  $\mu_{b_k} = \mu_{k_1}, \mu_{k_2}, \dots, \mu_{k_D}$ 는 최대 유사도 추정(maximum likelihood estimation)에 의하여 다음과 같이 구해줄 수 있다.

$$\begin{aligned}
 \mu'_{b_{k,i}} &= \frac{\sum_{i=1}^T \sum_{(n,m) \in \Omega_k} \gamma_t(n,m) \frac{y_{t,i} - \mu_{n,m,i}}{\sigma_{n,m,i}^2}}{\sum_{i=1}^T \sum_{(n,m) \in \Omega_k} \frac{\gamma_t(n,m)}{\sigma_{n,m,i}^2}} & (2) \\
 &i = 1, \dots, D, \quad k = 1, \dots, K
 \end{aligned}$$

$$\gamma_t(n,m) = \begin{cases} \frac{w_{n,m} N[y_t; \mu_{n,m}, C_{n,m}]}{\sum_{j=1}^M w_{n,j} N[y_t; \mu_{n,j}, C_{n,j}]} & \text{if } \hat{s} = n \\ 0 & \text{otherwise} \end{cases} & (3)$$

여기서  $D$ 는 특징 벡터의 차수이고  $\mu_{n,m}, C_{n,m}$ 은 상태  $n$ 에서  $m$ 번째 혼합성분에 해당하는 평균과 분산 벡터이고,  $w_{n,m}$ 은 상태  $n$ 에서  $m$ 번째 복합성분의 확률이고,  $N$ 은 노말 분포(normal distribution)이고  $s$ 는 입력 음성에 해당하는 상태열이다. 만일  $\mu_{n,m}$ 이  $k$ 번째 군 또는 코드워드에 군집되면  $(n,m) \in \Omega_k$ 이다.

변환 벡터에 의하여 변환된 화자독립 HMM에서 각각의

혼합성분의 평균  $\hat{\mu}_{n,m}$ 은 화자독립 HMM의 혼합성분의 평균  $\mu_{n,m}$ 에 변환 벡터를 더하여 다음과 같이 구해진다.

$$\hat{\mu}_{n,m} = \mu_{n,m} + \sum_{k=1}^K \mu_{b_k} I_{\Omega_k}(n,m), \quad \forall (n,m) \quad (4)$$

여기서  $I_{\Omega_k}(\cdot)$ 는 집합  $\Omega_k$ 를 나타내는 기호이다.

제안된 방법은 학습 과정에서 다음과 같이 세단계로 구현된다.

1. 화자독립 음소 HMM을 이용하여 다수의 입력 음성에 대한 초기 음소 열과 상태 분할 정보를 추정한다.
2. 변환 벡터는 식 (2)와 (3)을 이용하여 추정된다.
3. 화자독립 음소 모델들은 식 (4)를 사용하여 변환된다.
4. 단계 2, 3을 모델이 수렴될 때까지 반복하고 최종 변환 벡터와 최종 음소열을 인식 과정을 위하여 저장한다.

인식 과정에서 사용자는 이름을 말한다. 전화망에서는 발신자 번호 또는 사용자의 입력에 의하여 신원이 확인된다. 화자독립 음소 HMM이 해당 사용자의 변환 벡터를 사용하여 변환된 후에 인식기는 음소열과 변환된 화자독립 음소 HMM을 사용하여 입력 음성을 인식한다.

### 3. 실험 및 결과 고찰

#### 3.1 시스템 구성

실험에 사용된 데이터 베이스는 남성 5명과 여성 5명의 총 10명으로 구성하였다[6]. 각 화자는 15개의 단어를 발음하였다. 데이터 녹음은 전화선을 통하여 이루어 졌으며 각 화자는 각기 다른 환경에서 가급적 다른 종류의 전화기를 사용하여 몇 주 간격을 두고 녹음하였다. 음성 신호는 6.67kHz로 샘플링되었고 8bit  $\mu$ -law PCM으로 저장되었다. 학습에 사용된 데이터는 각 화자가 15개의 이름을 3회 반복한 것(15개×3회=45개/명)으로 구성하였으며, 인식에 사용된 데이터는 각기 다른 날짜에 수행한 5회의 녹음에서 각 화자가 15개의 이름을 10회 반복한 데이터(15개×10회=150개/명)로 구성하였다. 데이터 내용은 영어로 "Call office", "Call home" 등으로 구성되었다.

실험에 사용된 특징벡터는 12차 LPC 캡스트럼, 1차 차분 캡스트럼, 2차 차분 캡스트럼, 에너지, 1차 차분 에너지, 2차 차분 에너지의 총 39차 벡터로 구성되었다. 캡스트럼 계수는 30ms의 창 길이를 갖고 10ms씩 이동하면서 구한 10차 LPC 계수로부터 구하였다.

화자독립 음소 HMM은 연속음성 인식을 위하여 전화선을 통하여 녹음된 데이터베이스를 사용하여 학습된 모델을 사용하였다. 따라서 본 실험에 참여한 화자와 중복된 경우는 없었다. 이러한 모델은 각 음소마다 3개 또는 5개의 상태 수를 갖는 left-to-right 형태의 음소 모델 41개와 1개의 상태를 갖는 묵음 모델로 구성되었고 각각의 HMM은 연속밀도분포를 갖는 연속분포 HMM이다. 이러한 모델을 사용하여 입력 음성에 대한 음소 열을 추정하였다.

#### 3.2 기준 시스템의 성능 평가

제안된 방법의 비교 평가를 위하여 기준 시스템을 구성하

여 성능 평가를 수행하였다. 기준 시스템은 화자독립 음소 HMM을 사용한 화자 종속 음성 다이얼링 시스템으로 구현하였다. 구현된 시스템은 화자독립 음소모델을 사용하여 학습 데이터의 음소 열을 구하여 저장하고, 입력 음성을 인식할 때 저장된 음소 열과 화자독립 모델을 사용하였다. 이러한 방법은 저장 공간은 크게 줄일 수 있으나 화자독립 음소 HMM을 사용한 음소 열 추정 결과에 많은 오차가 발생하는 문제점이 있다. 이러한 오차를 줄이는 방법으로 묵음 사이의 무성음은 묵음으로 처리하는 등의 간단한 논리를 사용하여 음소 인식 오차를 줄일 수 있다[1-4].

본 논문에서는 이러한 오차를 줄이는 방법으로 음성 구간 검출 방법을 사용하였다. 즉 에너지 파라미터를 사용한 음성 구간 검출을 수행하여 음성으로 판단된 음성 구간의 음소 열만을 입력 음성에 대한 음소 열로 저장하였다. 표 1은 음성 구간 검출을 사용할 때와 사용하지 않을 때의 실험 결과를 보여준다. 표에서 음성 구간 검출을 사용한 경우는 잘못 인식된 음소 열을 제거하여 인식 오차가 4.2%에서 3.8%로 감소하였다.

표 1. 음성 구간 검출 사용여부에 따른 기준 시스템의 성능 비교.

Table 1. Performance Comparison of baseline system with or without endpoint detection.

	음성구간검출 미사용	음성구간검출 사용
인식 오차 (%)	4.2	3.8

#### 3.3 제안된 학습 알고리즘 성능평가

기존 화자독립 음소 HMM을 사용한 화자 종속 음성 다이얼링 시스템은 화자독립 모델을 음소 인식에 사용할 때 발생하는 오차로 인하여 화자종속 모델을 사용하는 방법보다는 인식 성능이 저하되는 문제점이 있다. 본 논문에서는 이러한 문제점을 개선하기 위하여 음소 열과 화자적응을 위한 모델 변환합수를 동시에 추정하는 방법을 제안하였다. 여기에서 화자 적응을 위해서 음소 군과 코드북 기반의 두 가지 종류의 확률적 매칭 방법을 사용하였다. 음소 군 기반의 확률적 매칭 방법에서는 변환 벡터의 수를 음소 군에 따라 결정하였다. 표 2는 음소 군에 따른 변환 벡터의 개수를 나타낸다.

표 2. 음소군 기반 확률적 매칭 방법에서 음소군에 따른 변환 벡터의 개수.

Table 2. The number of transformation vectors according to the phonetic class based stochastic matching.

변환 벡터 개수	음소 군
1	no phonetic class
2	silence and speech
3	silence, vowel and consonants
9	silence, vowel, diphthongs, semivowels, stops, . . .
14	14 silence, front, mid, back} vowel, diphthongs, liquids, glides, {voiced, unvoiced} stops, . . .
42	silence and all phones
180	all of SI HMM states to the acoustic phonetic class

위와 같은 데이터 베이스를 사용하고 화자적응 알고리즘을 사용한 음성 다이얼링 시스템의 성능은 그림 2와 같다. 그림 2(a)에서 가로축은 학습 과정에서의 음소 열과 변환 벡터 추정과정의 반복 횟수(iteration)를 나타낸다. 변환 벡터의 수는 음소의 형태에 따라서 1, 2, 3, 9, 14, 42, 180 개의 총 7가지 경우를 사용하였다. 그림에서 알 수 있듯이 변환 벡터가 1, 2, 3, 9의 경우에는 음소 열과 변환 벡터를 반복하여 추정하여도 인식 시스템의 성능이 기존 시스템보다 개선되지 않았다. 그러나 변환 벡터의 수를 14개 이상 사용하는 경우에는 시스템이 수렴하여 인식 오차가 감소하는 것을 알 수 있다. 여기서 변환 벡터의 수를 음소의 수와 같은 42개를 사용했을 때 가장 적은 인식 오차(2.3%)로 수렴하는 것을 알 수 있다

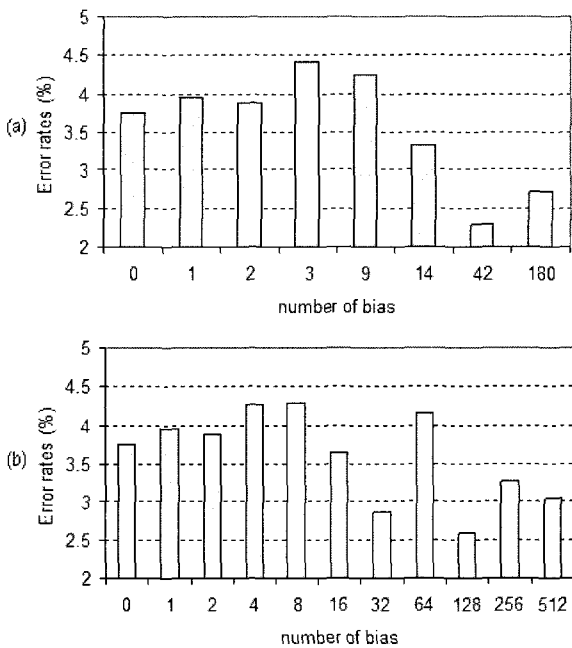


그림 2. 음소군(a)와 코드북(b)에 따른 제안된 학습 알고리즘을 사용한 음성 인식 시스템의 성능 평가.

Fig. 2. Performance of the speech recognition system using proposed algorithm according to the phonetic class(a) and codebook(b).

그림 2(b)에서는 변환 벡터를 모든 화자독립 음소 HMM의 평균 벡터를 집단화하여 적용하였다. 이때 사용된 코드북의 개수는 1, 2, 4, 8, 16, 32, 64, 128, 256, 512 개의 총 10가지 경우를 사용하였다. 그림에서 알 수 있듯이 변환 벡터가 8개 이하인 경우에는 음소 열과 변환 벡터를 반복하여 추정하여도 인식 시스템의 성능이 기존 시스템보다 개선되지 않았다. 그러나 변환 벡터의 수를 16개 이상 사용하는 경우에는 시스템이 수렴하기 시작하여 변환 벡터의 수가 128인 경우 가장 적은 인식 오차(2.6%)로 수렴하는 것을 알 수 있다. 그러나 이것은 음소 별로 변환 벡터를 사용한 경우의 최소 오차 2.3% 보다는 약간 높다.

본 실험에서는 제안된 방법의 성능을 기존의 방법과 비교하기 위하여 위에서 구현한 기존 시스템 이외에 다음과 같은 시스템을 구현하여 그 성능을 비교하였다.

표 3에서 기존 시스템은 음소 HMM과 음성 구간 검출을

사용하여 얻어진 음소 열을 사용한 시스템의 성능으로 3.8%의 인식 오차를 나타내었다. 두 번째는 기존 시스템에 변환 벡터 추정을 추가한 시스템의 인식 성능을 평가하였다. 이것은 본 논문에서 제안한 음소 열과 변환 벡터를 순환적으로 추정하는 방법과 성능 비교를 하려는 것이다. 변환 벡터만을 추정한 경우에도 인식 오차는 3.3%로 감소하는 것을 알 수 있다. 다음은 제안된 방법으로 음소 열과 변환 벡터를 순환적으로 추정한 방법의 결과이다. 인식 오차는 2.3%로 기존 시스템의 인식 오차가 1.9% 감소되었다. 마지막 열은 제안된 방법과 비교를 위하여 화자종속 HMM을 사용한 단독음 인식 시스템의 성능을 나타내었다. 이 경우에 인식 성능은 1.8%로 가장 높게 나타나지만 각 단어마다 모델을 저장하여야 하기 때문에 많은 저장 공간이 필요하다.

표 3. 제안된 학습 알고리즘과 기존 방법의 성능 비교.

Table 3. Performance comparison of the proposed training algorithm with convention methods.

시스템 형태	인식 오차(%)	메모리 (kbyte/user)
기존시스템	4.2	1.6
기존시스템과 음성구간검출	3.8	1.5
변환 벡터만 추정	3.3	1.5
제안된 시스템	2.3	1.5
이상적인 경우	2.3	0.7
화자종속 시스템	1.8	112

총 10명의 화자가 각 화자마다 15개의 이름을 3회 반복한 것(15개×3회=45개/명)을 학습에 사용한 경우, 음소 열과 화자독립 HMM을 이용한 시스템은 화자마다 평균 1.5Kbyte의 저장 공간이 필요한 반면 화자종속 HMM을 이용한 경우에는 화자마다 평균 112Kbyte가 필요하였다.

#### 4. 결 론

본 논문에서는 화자독립 음소모델을 사용한 음성 다이얼링 시스템의 성능을 개선하기 위하여 음소 열과 화자적응을 위한 모델 변환함수를 동시에 추정하는 방법을 제안하였다. 제안된 방법은 학습과정에서 학습 데이터의 음소 열과 화자적응을 위한 변환 벡터를 동시에 추정한 후 음소 열과 함께 저장하고, 인식 시에 화자독립 음소 HMM을 각 화자의 변환 벡터를 사용하여 변환한 후 인식을 수행하였다. 여기서 화자적응을 위한 변환 벡터는 확률적 매칭을 위한 최고 유사도 방법을 이용하였으며 음소 열과 함께 반복적으로 추정되었다. 본 논문에서는 음소 군(phonetic class)과 코드북(codebook) 기반의 두 가지 종류의 확률적 매칭 방법을 사용하여 변환 벡터를 추정하였다. 음소 군 기반 확률적 매칭 방법에서는 음소 군에 따라 다수의 변환 벡터가 사용되었다. 변환 벡터의 개수는 음향학적 음소 군에 따라 결정되었다. 코드북 기반의 확률적 매칭 방법에서는 모델 파라미터를 공통으로 사용하는 묶음(tying)의 개념을 사용하여 변환벡터의 개수를 결정하였다.

전환선을 통하여 구성된 데이터 베이스를 사용한 인식 실

험에서 기존 시스템의 인식오차 4.2%가 제안된 화자적응 방법을 사용하여 2.3%로 감소하여 약 45%정도의 인식 오차가 감소하는 것을 확인하였다.

저 자 소 개

참 고 문 헌

[1] Jain, N., Cole, R. Barnard, E., "Creating Speaker Specific Phonetic Templates with a Speaker-Independent Phonetic Recognizer: Implications for Voice Dialing", *Proc.of ICASSP*, pp. 881884, 1996

[2] Fontaine, V., Boulard, H., "Speaker-Dependent Speech Recognition Based on Phone-Like Units Models-Application to Voice Dialing", *Proc. of ICASSP*, pp. 15271530, 1997

[3] Ramabhadran, B., Bahl, L.R., deSouza, P.V., Padmanabhan, M., "Acoustic-Only Based Automatic Phonetic Baseform Generation", *Proc. of ICASSP*, pp. 22752278, 1998

[4] Deligne, S., Mangu, L., "On the use of Lattices for Automatic Generation of Pronunciations", *Proc. of ICASSP*, pp. 204207, 2003

[5] Sankar, A., Lee, C.H., "A Maximum-Likelihood Approach to Stochastic Matching for Robust Speech Recognition", *IEEE Trans. on Speech and Audio Processing*, Vol. 4., pp. 190202, 1996

[6] Sukkar, R.A., Lee, C.H., "Vocabulary Independent Discriminative Utterance Verification for Non-keyword Rejection in Subword based Speech Recognition", *IEEE Trans. on Speech and Audio Processing*, Vol. 4., pp. 420429, 1996



김원구(Weon-Goo Kim)  
 1987년 2월 : 연세대 전자공학과 학사  
 1989년 8월 : 연세대 전자공학과 석사  
 1994년 2월 : 연세대 전자공학과 박사  
 1994년 9월~현재 : 군산대 전자정보공학 부 교수  
 1998년 9월~1999년 9월 : Bell Lab, Lucent Technologies(USA) 객원연구원

관심분야 : 음성 신호처리, 음성 인식, 감성 인식, 음성 변환  
 Phone : 063) 469-4745  
 Fax : 063) 469-4699  
 E-mail : wgkim@kunsan.ac.kr