

## 외국어 발음오류 검출 음성인식기를 위한 MCE 학습 알고리즘\*

MCE Training Algorithm for a Speech Recognizer  
Detecting Mispronunciation of a Foreign Language

배 민 영\*\* · 정 용 주\*\*\* · 권 철 홍\*\*

Min Young Bae · Yong Joo Chung · Chul Hong Kwon

### ABSTRACT

Model parameters in HMM based speech recognition systems are normally estimated using Maximum Likelihood Estimation(MLE). The MLE method is based mainly on the principle of statistical data fitting in terms of increasing the HMM likelihood. The optimality of this training criterion is conditioned on the availability of infinite amount of training data and the correct choice of model. However, in practice, neither of these conditions is satisfied. In this paper, we propose a training algorithm, MCE(Minimum Classification Error), to improve the performance of a speech recognizer detecting mispronunciation of a foreign language. During the conventional MLE(Maximum Likelihood Estimation) training, the model parameters are adjusted to increase the likelihood of the word strings corresponding to the training utterances without taking account of the probability of other possible word strings. In contrast to MLE, the MCE training scheme takes account of possible competing word hypotheses and tries to reduce the probability of incorrect hypotheses. The discriminant training method using MCE shows better recognition results than the MLE method does.

**Keywords :** Detection of mispronunciation, Discriminant, MCE method

### 1. 서 론

최근 국제화 시대에 있어서 외국어 발화의 발음 교정에 대한 관심이 높아지고 있다. 이러한 발화 양상의 교정은 극소수의 언어·음성학 전문가들에 의해서만 수행되고 있는 형편이며, 아동의 언어발달과 성인의 외국어 습득을 위한 발음교정은 여러 차원에서 교육 및 재교육이 필요하나, 비용과 접근의 불편함 때문에 그 해결의 중요성에도 불구하고 기회 제공이 원활히 이루어져 오고 있지 않다. 이를 위해서 음성인식 기술을 이용하여 발음상의 미비함, 조음기관 동작의 부정확성을 정확히 인식하고 그 각각의 범주에 따른 교정방안을 컴퓨터를 통하여 제시하는 시스템의 개발이 필요

\* 본 연구는 한국과학재단 목적기초연구(R01-2002-000-00283-0) 지원으로 수행되었음.

\*\* 대전대학교 정보통신공학과

\*\*\* 계명대학교 전자공학과

한 실정이다.

이와 같은 필요성에 의해 음성인식 엔진을 응용한 외국어 발음 학습기가 출시되고 있으나, 출시된 제품은 외국에서 개발된 엔진이거나 엔진 활용용도가 일반적인 상업용 엔진을 사용하기 때문에 한국인에게 적합한 수준의 학습 결과를 제시하지 못하고 있다. 또한 기존의 외국어 발음 학습 시스템은 일반적인 음성 인식을 기반으로 하여 학습자의 발화 데이터로부터 출력되는 음향모델과의 음향 유사성(acoustic likelihood)들을 백분율 점수로 나타내어 발음의 성취도를 평가하는 수준이다. 이러한 시스템은 잘못 발음된 단어 내 음소를 검출하여 정확한 교정 정보를 학습자에게 제공할 수 없다. 기업의 상품이 아닌 학술적 연구로서 외국어 발화 오류 검출 음성인식기에 대한 연구는 그리 많지 않다. 그나마 문장 단위의 유창성을 측정하는 발성 평가를 수행하는 음성인식기에 대한 연구가 주를 이루고 있다[1][2]. 본 연구에서는 단어 단위의 발화를 토대로 음소 단위로 오류 발음을 검출하는 음성인식기를 연구한다.

본 논문에서는 음성인식 기술의 응용분야인 외국어 학습기에 초점을 맞추어 한국인의 외국어 발음 교정시스템을 위한 발음오류유형 자동분류 음성인식기를 다룬다. 특히 본 논문에서는, 발음에 따른 오류 양상을 음성인식기로 분류하여 그 오류에 해당하는 교정지침을 제공하기 위하여, 원어민의 발음과 그 오류에 해당하는 유사발음을 정확하게 구분해 낼 수 있도록 음성인식기의 성능을 향상시키는 데 목적이 있다. 음성인식기의 인식결과는 전문음성학자의 청취판단과 상관관계가 높아야 하므로 음성학자의 청취판단과의 결과를 비교한다. 즉, 본 연구의 초점은 발음오류 검출 음성인식기가 인식해 낸 결과와 해당 언어 음성학 전문가의 청취판단 결과가 가능한 한 유사하게 나타내게 하는 데 있다.

HMM은 음성의 시간적 특징을 비교적 잘 반영하는 뛰어난 특성으로 인해 음성인식 분야에서 성공적으로 적용되어 왔다. HMM은 동적인 발화 패턴을 처리하기 위해 효율적인 상태 천이와 통계 모델을 사용하여 음성의 특징을 연속적이며 정적으로 표현한다. 이 HMM은 MLE(Maximum Likelihood Estimation)를 기반으로 훈련한다. MLE 학습 방법은 주어진 모델의 데이터에 대한 likelihood을 최대화 시키는 것이다. 이 MLE는 방대한 양의 음성 데이터와 모델의 올바른 선택이라는 조건들이 만족되면 에러의 확률을 최소화 시킨다는 점에선 최적의 방법이다[3]. 그러나 이 조건들은 현실 상황에서는 거의 만족되기 어려운 것들이다. 본 논문에서는 이런 HMM의 단점을 보완하여 인식률을 향상시키는 방법으로 MCE(Minimum Classification Error) 알고리즘을 이용한 변별적 학습방법을 제안한다.

본 논문의 구성은, 서론에 이어 2장에서 제안한 MCE 학습 방식을 설명하고, 3장에서 실험 방법 및 결과를 기술하고, 그리고 4장에서 결론을 맺는다.

## 2. MCE 학습 알고리즘

MCE는 인식에러를 최소화하는 학습방법이다. 이는 최소인식에러 확률 값을 정확히 구하는 대신 인식 에러에 발생하는 비용함수를 최소화하는 방법이다. 그림 1은 MCE 학습 방법의 전체 개요를 보여 준다. 수집한 음성DB에서 특징계수를 추출하고 기존 MLE 방식으로 생성된 HMM 모델을

이용하여 비터비 디코딩을 수행하여 인식결과를 구한다. 인식결과에서 오인식 정도를 추정하고 MCE 학습 방법을 이용하여 루프를 순환하여 만들어진 새로운 파라미터 값으로 보정된 HMM 모델로 최종인식을 수행한다.

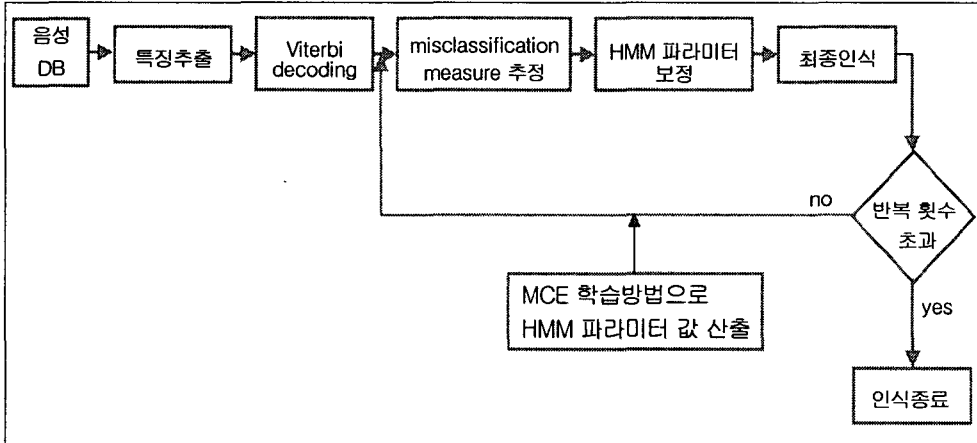


그림 1. MCE 훈련 방식

MCE 방식에서는 각 모델(class)  $A_i, i = 1, \dots, L$ 에 대한 변별함수  $g_i(X, A_i)$ 가 필요하다.  $L$ 은 모델의 수,  $A_i$ 는 HMM 파라미터 값을 나타낸다. 가장 큰 변별함수 값을 갖는 모델을 선택함으로써 입력  $X$ 가 결정된다. 여기에서는 비터비 디코딩에서 구한 log-likelihood 값이 변별함수로 사용된다.

본 논문에서는 음소 모델 HMM이 인식의 기본적인 단위로서 사용되고, 위에서 언급된  $A_i (i = 1, \dots, L)$  모델은 음소 모델의 연쇄로 구성된 단어 모델 HMM에 상응된다. 그리고 최소화 하고자 하는 최종 함수는 단어 인식 에러 개수이다.

$j$ 번째 상태에서 관측벡터  $x$ 의 출력확률 밀도함수는 식 (1)과 같다.

$$b_j(x) = \sum_{m=1}^M c_{jm} N(x, \mu_{jm}, \Sigma_{jm}) \tag{1}$$

여기에서  $M$ 은 상태(state)의 mixture 개수이다.  $c_{jm}$ 은 mixture weight이고  $\sum_{m=1}^M c_{jm} = 1$ 을 만족한다.  $N(x, \mu_{jm}, \Sigma_{jm})$ 은 평균 벡터  $\mu_{jm}$ 과 대각선의(diagonal) 공분산 행렬  $\Sigma_{jm}$ 으로 나타낸다.

$i$  번째 단어 모델  $A_i$ 에서 최적의 경로에 따른 입력 벡터  $X = \{x_0, \dots, x_T\}$ 의 log-likelihood 값 즉 변별함수는 식 (2)와 같이 나타낼 수 있다.

$$g_i(X, \Lambda_i) = \log b_{s_0}^i(x_0) + \sum_{t=1}^T \log a_{s_{t-1}s_t}^i + \sum_{t=1}^T \log b_{s_t}^i(x_t). \quad (2)$$

$S = \{s_0, \dots, s_T\}$ 는 최적경로에 따른 상태열을 나타낸다.  $T$ 는 입력 단어  $X$ 의 프레임 수이다.  $\log b_{s_t}^i(x_t)$ 는  $i$  번째 모델의 상태  $s_t$ 에서 관측벡터  $x_t$ 의 출력확률의 로그 값이다.  $a_{s_{t-1}s_t}^i$ 는  $i$  번째 모델에서 상태  $s_{t-1}$ 에서  $s_t$ 로의 천이확률을 나타낸다.  $c = \arg \max_k g_k(X, \Lambda_k)$ 를 만족하면, 인식기는 입력  $X$ 를  $c$  번째 단어로 인식한다.

식 (3)은 오인식 정도를 나타내는 함수이다.

$$d_c(X) = -g_c(X, \Lambda_c) + \left[ \frac{1}{M-1} \sum_{i \neq c} g_i(X, \Lambda_i) \right]^{\frac{1}{\eta}} \quad (3)$$

$\eta$ 은 양수,  $M$ 은 전체 단어의 개수,  $X$ 는  $c$  번째 모델에 대한 입력 단어이다. 우측 첫째 항은 입력 단어에 대한 log-likelihood 값이고, 둘째 항은 다른 단어들에 대한 log-likelihood 값의 합을 나타낸다. 이 식에서 큰 양수 값  $d_c(X)$ 은 오인식이 발생했다는 것을 의미하고, 작은 양수 값이나 음수 값은 오인식이 발생하지 않았음을 나타내므로, 이 식은 오인식 정도를 보여주는 함수라고 말할 수 있다.

다음으로, 최소 에러에 대한 비용함수를 정의한다. 식 (4)는 모델  $i$ 에 대한 비용함수이다.

$$l_i(d_i(X)) = \frac{1}{1 + \exp(-\gamma(d_i(X)))} \quad (4)$$

여기에서  $\gamma$  값은 고정된 값으로 비용함수의 기울기를 조절한다. 이 비용함수의 최소화는 에러 개수의 최소화로 직접 연결된다.

GPD(Generalized Probabilistic Descent) 방식이 효과적으로 HMM 파라미터 값을 반복적으로 조절하기 위해 사용된다[4]. 이 방법에서 입력 단어  $X$ 에 대하여 모델 파라미터  $\Lambda$ 는 다음 식 (5)에 의해 구할 수 있다.

$$\Lambda_{n+1} = \Lambda_n - \epsilon_n U \nabla l_c(d_c(X)) \quad (5)$$

$U$ 는 양수의 한정된 행렬이고,  $\{\epsilon_n\}$ 은  $\sum_{n=1}^{\infty} \epsilon_n \rightarrow \infty$ ,  $\sum_{n=1}^{\infty} \epsilon_n^2 < \infty$ 을 만족한다. HMM 파라미터를

조절하기 위한 이 기울기(gradient) 방법에서 mixture weight는  $\sum_m c_{jm} = 1$ ,  $c_{jm} \geq 0$ 를, 천이확

들은  $\sum_i a_{ij} = 1$ ,  $a_{ij} \geq 0$ 를, 분산은 양수 값을 가져야 한다는 조건을 만족해야 한다.

먼저 파라미터들의 대수를 구하고 gradient descent에 의해 값을 조절한다. 그 다음에 조절된 파라미터들의 역지수를 구하고 마지막으로 위 조건을 만족시키기 위해 값을 normalize 한다[5]. 식 (2), (3), (4)로부터 음소 p에 대한 HMM 파라미터에 대한  $l_c(d_c(X))$ 의 미분계수는 식 (6), (7), (8), (9)와 같이 구할 수 있다.

$$\frac{\partial l_c(d_c(X))}{\partial a_{ij}^p} = k \cdot \left( - \sum_{t=1}^T \text{sgn}(a_{ij}^p, a_{s_{t-1}, s_t}^c) + \frac{1}{M-1} \sum_{i \neq c} \sum_{t=1}^T \text{sgn}(a_{ij}^p, a_{s_{t-1}, s_t}^i) \right),$$

$$a_{ij}^p = \log a_{ij}^p \quad (6)$$

$$\frac{\partial l_c(d_c(X))}{\partial c_{jm}^p} = k \cdot \gamma_{jm}^p \cdot \left( - \sum_{t=1}^T \text{sgn}(c_{jm}^p, c_{s_t, m}^c) + \frac{1}{M-1} \sum_{i \neq ct=1}^T \sum_{t=1}^T \text{sgn}(c_{jm}^p, c_{s_t, m}^i) \right),$$

$$c_{jm}^p = \log c_{jm}^p \quad (7)$$

$$\frac{\partial l_c(d_c(X))}{\partial \sigma_{jmd}^p} = k \cdot \gamma_{jm}^p \cdot \left( - \sum_{t=1}^T \text{sgn}(\sigma_{jmd}^p, \sigma_{s_t, md}^c) \right)$$

$$+ \frac{1}{M-1} \sum_{i \neq ct=1}^T \sum_{t=1}^T \text{sgn}(\sigma_{jmd}^p, \sigma_{s_t, md}^i) \cdot \left( -\frac{1}{2} + \frac{1}{2} \cdot \frac{(x_{td} - \mu_{jmd}^p)^2}{\sigma_{jmd}^p} \right),$$

$$\sigma_{jmd}^p = \log \sigma_{jmd}^p \quad (8)$$

$$\frac{\partial l_c(d_c(X))}{\partial \mu_{jmd}^p} = k \cdot \gamma_{jm}^p \cdot \left( - \sum_{t=1}^T \text{sgn}(\mu_{jmd}^p, \mu_{imd}^c) \right)$$

$$+ \frac{1}{M-1} \sum_{i \neq ct=1}^T \sum_{t=1}^T \text{sgn}(\mu_{jmd}^p, \mu_{jmd}^i) \cdot \left( \frac{x_{td} - \mu_{jmd}^p}{\sigma_{jmd}^p} \right) \quad (9)$$

여기에서

$$k = \frac{\exp^{\gamma \cdot d_c} \cdot \gamma}{(1 + \exp^{-\gamma \cdot d_c})^2}, \quad (10)$$

$$\gamma_{jm}^p = \frac{c_{jm}^p N(x_t, \mu_{jm}^p, \sum_{jm}^p)}{b_j^p(x_t)} \quad (11)$$

위 식에서  $\sigma_{jmd}^p$ 는  $\sum_{jm}^p$ 의 d 번째 대각선 값이고,  $x_{td}$ 는  $x_t$ 의 d 번째 값이다.  $sgn(A, B)$ 는  $A, B$ 가 같은 HMM의 파라미터에 대응하면 1로, 그렇지 않으면 0으로 정의된다.

### 3. 실험 방법 및 결과

#### 3.1 실험 방법

실험을 위한 발음 네트워크는 그림 2에서 보는 바와 같이 일본어 음소 [k]에 대하여 유사발음 일본어 음소(ky, g, gy)와 한국어 오류음소(/ㅋ/khc, /ㄱ/kkc, /무성 ㄱ/kc, /유성 ㄱ/gc)를 통합한 네트워크를 사용하였다. 또한, 일본어 음소 [p]에 대하여 유사발음 일본어 음소(py, b, by)와 한국어 오류음소(/ㅍ/phc, /ㅂ/ppc, /무성 ㅂ/pc, /유성 ㅂ/bc)를 통합한 네트워크도 사용하였다. 일본어 음소 모델[6]은 일본인 50 인이 일본어 PBW 216 개 단어를 발화한 일본어 음성DB로, 한국어 음소 모델은 한국인 50 인이 한국어 PBW 452 개 단어를 발화한 한국어 음성DB로 생성하였다. 특징벡터는 MFCC 39 차를 사용하였고, HMM 모델은 3 state left-to-right model을 사용하여, Mixture 수는 일본어, 한국어 음소모델 모두 15로 하였다.

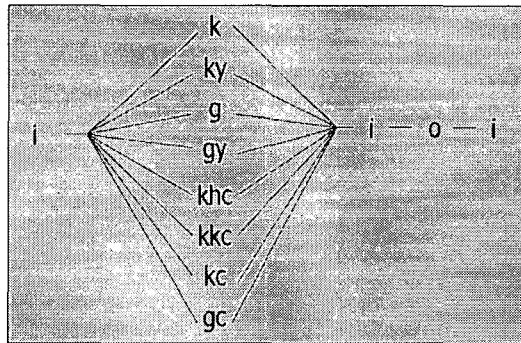


그림 2. 오류발음 검출 발음 네트워크  
(일본어 단어 i-k-o-i 인 경우)

본 논문에서 구현한 외국어 발음오류 검출 음성인식기가 그림 3에 보인다. 먼저, 한국인이 발화한 일본어 음성DB와 원어민이 발화한 일본어 음성DB에서 각각 음성신호의 특징을 추출하여 MLE 방식을 이용하여 음소 단위로 훈련하여 각 언어의 음소별 HMM 모델을 구성한 후 그 모델들을 결합시킨다. 이 결합된 모델은 입력된 음성과 함께 인식에 사용된다. 그리고 본 논문에서 제안한 MCE 학습 방법을 이용하여 최적화된 파라미터 값을 계산하고 각 언어 음소별 HMM 파라미터 값(mean, variance, mixture weight, transition probability 등)을 보정(update)한다. 이런 식으로 루프를 순환하여 각 언어 모델에 대한 최적화된 HMM을 가지고 인식을 하게 된다.

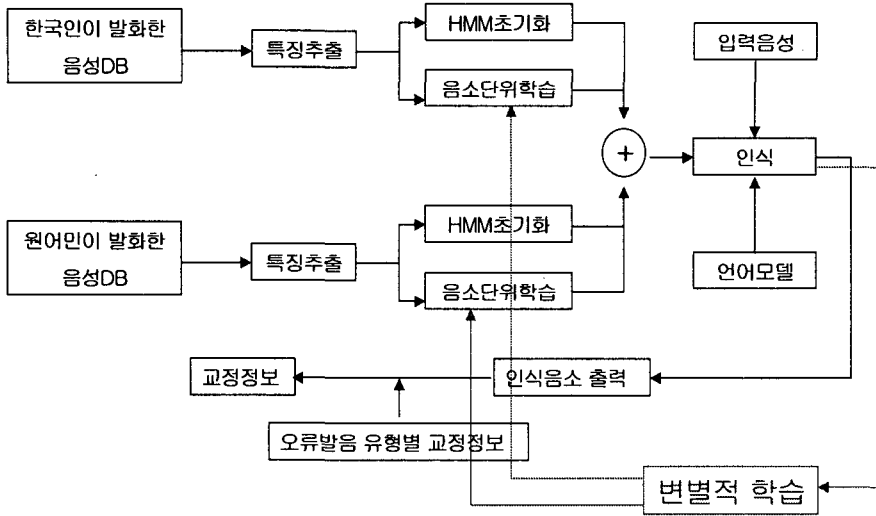


그림 3. 발음오류 검출 음성인식기

발음오류 자동분류 음성인식기의 성능을 검증하기 위하여 일본어 비전공 한국인 대학생 남성 15 명, 여성 15 명 등 30 명이 일본어 음소 [k]가 포함된 일본어 단어 24 개를 발화한 음성 데이터와, 일본어 비전공 한국인 대학생 남성 12 명, 여성 12 명 등 24 명이 일본어 음소 [p]가 포함된 일본어 단어 22 개를 발화한 음성 데이터를 수집하였다. 이들을 대상으로 인식 실험을 수행하여 유사 음소별 인식결과를 구했다.

### 3.2 기존 MLE 학습 방식을 이용한 실험 결과

이 절에서는 기존 MLE 학습 방식을 이용하여 실험한 결과를 보여 준다. 음성학자의 청취 판단 결과 어두 음소와 어중 음소에서 각 음소별로 크게 차이를 보이므로[7], 실험결과를 어두와 어중으로 나누어 분석하였다. 그리고 청취판단 결과에서 발화자들(일본어 비전공 한국인 대학생)이 일본어로 발화한 경우는 없다고 보았으므로 일본어 [k, ky, g, gy]와 [p, py, b, by]에 대한 실험결과는 제외하였다. 표 1, 2에 일본어 음소 [k]와 [p]에 대하여 청취판단과의 일치도가 보인다. 어중의 /ㄱ/와 어중의 /ㅍ/ 음소를 제외한 나머지 경우에 성능 개선의 필요성이 있음을 알 수 있다.

표 1. [k] 음소의 청취판단 결과와의 일치도 비교 (MLE 방식)

| 음소       | 청취판단 개수 |     |     | 청취판단과 일치한 인식 결과 (개수) |     |     | 일치 비율 (%) |      |      |
|----------|---------|-----|-----|----------------------|-----|-----|-----------|------|------|
|          | 어두      | 어중  | 계   | 어두                   | 어중  | 계   | 어두        | 어중   | 평균   |
| ㄱ(kkc)   | 4       | 318 | 322 | 1                    | 191 | 192 | 25.0      | 60.1 | 59.6 |
| ㅋ(khc)   | 67      | 130 | 197 | 13                   | 47  | 60  | 19.4      | 36.2 | 30.5 |
| ㄲ(kc/gc) | 139     | 62  | 201 | 37                   | 6   | 43  | 26.6      | 9.7  | 21.4 |

표 2. [p] 음소의 청취판단 결과와의 일치도 비교 (MLE 방식)

| 음소       | 청취판단 개수 |     |     | 청취판단과 일치한 인식 결과 (개수) |     |     | 일치 비율 (%) |      |      |
|----------|---------|-----|-----|----------------------|-----|-----|-----------|------|------|
|          | 어두      | 어중  | 계   | 어두                   | 어중  | 계   | 어두        | 어중   | 평균   |
| ㅃ(ppc)   | 25      | 109 | 134 | 6                    | 40  | 46  | 24.0      | 36.7 | 34.3 |
| ㅍ(phc)   | 133     | 223 | 356 | 48                   | 122 | 170 | 36.1      | 54.7 | 47.8 |
| ㅂ(pc/bc) | 10      | 28  | 38  | 0                    | 0   | 0   | 0.0       | 0.0  | 0.0  |

## 3.3 제안한 MCE 학습 방식을 이용한 실험 결과

표 3과 4는 MCE 학습 방식을 이용한 인식기의 결과와 청취판단과의 일치도를 보여 준다. 표 3을 보면 표 1과 비교해서 어두의 /ㄱ/을 제외한 모든 경우에서, 표 4를 보면 표 2와 비교해서 어두, 어중의 /ㄷ/을 제외한 모든 경우에서 청취판단과의 일치도가 증가함을 알 수 있다. [k] 음소인 경우 표 1과 3에서 음소별로 살펴보면, /ㄱ/ 음소인 경우 59.6%에서 65.5%로 5.9%, /ㅋ/ 음소는 30.5%에서 41.6%로 11.1%, /ㄷ/ 음소는 21.4%에서 35.3%로 13.9%가 향상되었다. 청취판단과의 비교대상 전체 발화 데이터 720 개 중에서, MLE 방식은 41.0%(295 개)의 일치도를 보여 주었고, MCE 방식은 50.6%(364 개)의 일치도를 보여 주어 9.6%의 성능 개선을 이루었다. 그리고 [p] 음소인 경우 표 2와 4에서 음소별로 살펴보면, /ㅃ/ 음소인 경우 34.3%에서 40.3%로 6.0%, /ㅍ/ 음소는 47.8%에서 59.6%로 11.8%가 향상되었다. 청취판단과의 비교대상 전체 발화 데이터 528 개 중에서, MLE 방식은 40.9%(216 개)의 일치도를 보여 주었고, MCE 방식은 50.4%(266 개)의 일치도를 보여 주어 9.5%의 성능 개선을 이루었다. 이와 같은 결과로부터 MCE 방식이 상당한 성과를 얻었음을 알 수 있다.

표 3. [k] 음소의 청취판단 결과와의 일치도 비교 (MCE 방식)

| 음소       | 청취판단 개수 |     |     | 청취판단과 일치한 인식 결과 (개수) |     |     | 일치 비율 (%) |      |      |
|----------|---------|-----|-----|----------------------|-----|-----|-----------|------|------|
|          | 어두      | 어중  | 계   | 어두                   | 어중  | 계   | 어두        | 어중   | 평균   |
| ㄱ(kkc)   | 4       | 318 | 322 | 1                    | 210 | 211 | 25.0      | 66.0 | 65.5 |
| ㅋ(khc)   | 67      | 130 | 197 | 18                   | 64  | 82  | 26.9      | 49.2 | 41.6 |
| ㄷ(kc/gc) | 139     | 62  | 201 | 61                   | 10  | 71  | 43.9      | 16.1 | 35.3 |

표 4. [p] 음소의 청취판단 결과와의 일치도 비교 (MCE 방식)

| 음소       | 청취판단 개수 |     |     | 청취판단과 일치한 인식 결과 (개수) |     |     | 일치 비율 (%) |      |      |
|----------|---------|-----|-----|----------------------|-----|-----|-----------|------|------|
|          | 어두      | 어중  | 계   | 어두                   | 어중  | 계   | 어두        | 어중   | 평균   |
| ㅃ(ppc)   | 25      | 109 | 134 | 11                   | 43  | 54  | 44.0      | 39.4 | 40.3 |
| ㅍ(phc)   | 133     | 223 | 356 | 63                   | 149 | 212 | 47.4      | 66.8 | 59.6 |
| ㅂ(pc/bc) | 10      | 28  | 38  | 0                    | 0   | 0   | 0.0       | 0.0  | 0.0  |



#### 4. 결 론

기존 음성인식 모델은 HMM 기반의 MLE에 의한 훈련 방법을 이용하여 생성된다. MLE 방식은 HMM log-likelihood 증가를 목표로 하는 data fitting에 초점을 두고 있다. 그러나 이 MLE는 충분한 양의 음성데이터와 음성발성에 대한 모델의 정확성이 요구된다. 이 요구 조건들이 충족되지 않는다면 MLE가 인식에러를 최소화 한다고 보장하기 어렵다. 그러나 이런 요구 조건들은 현실 상황에서는 충족되기 어렵다.

따라서 본 논문에서는 인식기의 최소 인식에러를 나타내기 위해 변별학습 기반의 MCE 알고리즘을 통하여 인식기의 정확성을 높이고자 했다. 기존의 MLE 방식은 특정 한 음소의 유사도 값을 증가시키는데 목적을 두었으나, MCE 알고리즘은 인식에러를 최소화하는 학습방법으로 이는 최소 인식에러 확률 값을 정확히 구하는 대신 인식 에러에 발생하는 비용함수를 최소화하는 방법이라고 할 수 있다. 이런 MCE 알고리즘과 같은 변별적인 학습 방법을 통하여 본 논문에서는 음성인식의 인식률이 증가함을 볼 수 있었다.

앞으로는 MMI(Maximum Mutual Information)와 같은 다른 변별적 학습 알고리즘을 통하여 발음오류 검출 음성인식기의 성능을 개선할 계획이다.

#### 참 고 문 헌

- [1] 박전규, 이준조, 김영창, 허용수, 이석재, 이종현. 2003. "발성 평가를 위한 영어 음성인식기의 개발." 대한음성학회 2003 가을 학술대회, pp. 37-40.
- [2] 이석재, 박전규. 2003. "한국인의 영어 문장 발음에 대한 한국인/원어민/ILT 평가 점수 사이의 상관관계." 대한음성학회 2003 가을 학술대회, pp. 83-87.
- [3] Woodland, P. C., Povey, D. 2000. "Large scale discriminative training for speech recognition." Proc. of ISCA ITRW ASR 2000, pp. 7-16, Paris.
- [4] Chou, W., Juang, B. H., Lee, C. H. 1992. "Segmental GPD training of HMM based speech recognizer", Pprod. of ICASSP 92, pp. 473-476, San Francisco.
- [5] Chung, Y. J., Un, C. K. 1996. "Multilayer perceptrons for state-dependent weightings of HMM likelihoods." *Speech Communication*, Vol. 18, pp. 79-89.
- [6] Kawahara, T., Kobayashi, T. et al. 1998. "Sharable software repository for Japanese large vocabulary continuous speech recognition." Proc. ICSLP 98, pp. 3257-3260, Sydney.
- [7] 배민영, 권철홍. 2004. "한국인의 외국어 발화오류검출 음성인식기에서 청취판단과 상관관계가 높은 기계 스코어링 기법." *음성과학*, 11권 2호, pp. 217-226.

접수일자: 2004. 11. 01

게재결정: 2004. 11. 28

## ▲ 배민영

대전 동구 용운동 96-3 (우: 300-716)  
대전대학교 정보통신공학과 대학원  
Tel: +82-42-280-2567, Fax: +82-42-284-0109  
E-mail: missbea79@hotmail.com

## ▲ 정용주

대구 달서구 신당동 1000 (우: 704-701)  
계명대학교 전자공학과  
Tel: +82-53-580-5925, Fax: +82-53-580-5929  
E-mail: yjjung@kmu.ac.kr

## ▲ 권철홍: 책임저자

대전 동구 용운동 96-3 (우: 300-716)  
대전대학교 정보통신공학과  
Tel: +82-42-280-2555, Fax: +82-42-284-0109  
E-mail: chkwon@dju.ac.kr