

# 타언어권 화자 음성인식을 위한 혼잡도에 기반한 다중발음사전의 최적화 기법\*

김민아(LG전자), 오유리(GIST), 김홍국(GIST),  
이연우(목포대), 조성의(목포대), 이성로(목포대)

## <차 례>

- |                                   |                             |
|-----------------------------------|-----------------------------|
| 1. 서론                             | 4. 제안된 발음사전 최적화 기법          |
| 2. 음성 코퍼스와 Baseline 음성인식기         | 4.1. 혼잡도 측정 방법              |
| 2.1. 음성 코퍼스                       | 4.2. 혼잡도 측정을 통한 발음사전<br>최적화 |
| 2.2. Baseline 음성인식기               | 5. 실험 및 결과                  |
| 3. 간접적 데이터 기반 타언어권 화자<br>발음변이 모델링 | 6. 결론                       |

## <Abstract>

### Optimizing Multiple Pronunciation Dictionary Based on a Confusability Measure for Non-native Speech Recognition

Mina Kim, Yoo Rhee Oh, Hong Kook Kim,  
Yeonwoo Lee, Sung Eui Lee, Seong Ro Lee

In this paper, we propose a method for optimizing a multiple pronunciation dictionary used for modeling pronunciation variations of non-native speech. The proposed method removes some confusable pronunciation variants in the dictionary, resulting in a reduced dictionary size and less decoding time for automatic speech recognition (ASR). To this end, a confusability measure is first defined based on the Levenshtein distance between two different pronunciation variants. Then, the number of phonemes for each pronunciation variant is incorporated into the confusability measure to compensate for ASR errors due to words of a shorter length. We investigate the effect of the proposed method on ASR performance, where Korean is selected as the target language and Korean utterances spoken by Chinese native speakers are considered as non-native speech. It is shown from the experiments that an ASR system using the multiple pronunciation dictionary optimized by the proposed method can provide a relative average word error rate reduction of 6.25%, with 11.67% less ASR decoding time, as compared with that using a multiple pronunciation dictionary without the optimization.

\* Keywords: Non-native speech recognition, Pronunciation model adaptation, Confusability reduction, Data-driven method.

\* 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2007-314-D00245).

## 1. 서 론

보다 쉽고 간편한 어플리케이션의 제공을 위하여 음성인식의 수요가 증가하고 있다. 뿐만 아니라, 세계화로 인해 국가 간에 교류가 빈번해지면서 화자가 타국어를 발화해야 하는 상황이 빈번하게 발생한다. 이와 같은 시대의 흐름에 따라, 타언어권 화자의 음성을 인식해야하는 경우가 종종 생긴다. 그러나 타언어권 화자의 경우, 모국어의 영향이나 유창하지 못한 발음으로 원어민의 발음과는 다르게 발음한다. 따라서 타언어권 화자의 음성인식은 원어민 화자의 음성인식에 비해 상대적으로 매우 낮은 성능을 보인다. 그러므로 효과적인 타언어권 화자 음성인식을 연구할 필요가 있다[1]. 타언어권 화자의 음성인식 성능을 향상시키기 위한 방법으로는 음향모델 적용, 발음모델 적용, 그리고 언어모델 적용 등 크게 세 가지 접근 방법으로 나눌 수 있다[2]. 본 논문에서는 타언어권 화자의 음성인식 성능을 향상시키기 위해 발음모델 적용에 초점을 맞추었다.

발음모델 적용 방법 중에서는 일반적으로 데이터 기반 발음모델 적용이 가장 효과적인 방법으로 알려져 있으며[3], 이는 간접적인 방법과 직접적인 방법으로 나눌 수 있다. 직접적인 데이터 기반 발음변이 모델링은 훈련 데이터로부터 추출한 발음변이를 음성인식에 사용되는 발음사전에 직접적으로 적용하는 기법이다. 그러나 발음변이를 추출한 훈련 데이터가 테스트 데이터의 단어와 상이한 경우, 그 성능이 크게 향상되지 않거나 오히려 저하되기도 한다. 반면에 간접적인 데이터 기반 발음변이 모델링은 훈련 데이터로부터 추출한 규칙을 사용하여 어떤 단어에도 적용할 수 있다.

간접적 데이터 기반 발음 변이 모델링을 이전 연구에서 제안된 바가 있다[4]. 이 방법은 타언어권 화자를 위한 음성인식에서 타언어권 화자의 발음 변이로부터 변이 규칙을 생성하고 이를 사전에 적용시키는 방법이다. 생성된 발음사전은 다중 발음사전으로 한 단어에 대한 발음열이 하나 또는 그 이상을 가지게 된다. 그러므로 사전의 크기가 증가하게 되어 음성을 인식할 때 많은 메모리와 계산량을 필요로 하게 된다. 또한, [5]에서 발음사전의 크기가 크게 증가하면, 혼잡도가 증가하고 이로 인해 인식 성능이 크게 향상되지 않거나 오히려 저하되는 현상을 보인다고 한다. 따라서 효과적으로 발음변이를 선택함으로써 발음사전의 혼잡도를 줄이고 인식 성능 향상에도 기여할 수 있다. Amdal 등은 발음사전을 줄이기 위해서 변이 규칙 확률과 로그 확률을 이용하는 방법이 연구하였다[6]. 유사하게, Hernandez-Abrego 등은 non-uniform pruning 또는 consensus based-pruning을 이용하여 발음사전의 크기를 줄이는 방법을 제안하였다[7]. 그러나 이러한 기존의 연구는 한 단어에 발생하는 다중 발음열에 대해 가지치기하여 발음사전의 크기를 줄임으로써 다른 단어와의 관계에서 발생하는 혼잡도는 줄이지 못했다. 발음사전에서 한 단어에 대한 다중 발음열이 다른 단어의 발음열과 유사하거나 같은 경우가 발생하는 것을

볼 수 있다.

본 논문에서는 사전의 크기를 최적화하여 음성인식에서의 계산량을 감소시키기 위해 혼잡도를 증가시키는 발음열을 제거하는 기법을 제안한다. 먼저, 서로 다른 두 발음열 사이의 Levenshtein 거리와 발음열의 길이를 고려한 혼잡도 측정 방법을 정의한다. Levenshtein 거리로 두 발음열 사이의 유사성을 측정할 수 있다. 또한 짧은 길이의 발음열일수록 더 많은 오인식을 발생시키므로 발음열의 길이를 함께 고려해 준다. 이렇게 정의된 혼잡도 측정 방법을 다중 발음사전의 각 변이 발음열에 적용하여 혼잡도를 증가시키는 발음열을 제거함으로써 발음사전을 최적화시켜 준다.

본 논문의 구성은 다음과 같다. 서론에 이어서 2장에서는 본 논문에서 사용된 음성 코퍼스와 baseline 음성인식 시스템을 설명하고, 3장에서 간접적 데이터 기반의 다중 발음사전 생성 방법을 설명한다. 본 논문에서 제안한 다중 발음사전의 혼잡도 감소를 위한 최적화 방법은 4장에서 기술한다. 5장에서는 제안된 방법을 이용한 음성인식 실험과 그 결과를 설명한다. 마지막으로 6장에서 본 논문의 결론을 맺도록 한다.

## 2. 음성 코퍼스와 Baseline 음성인식기

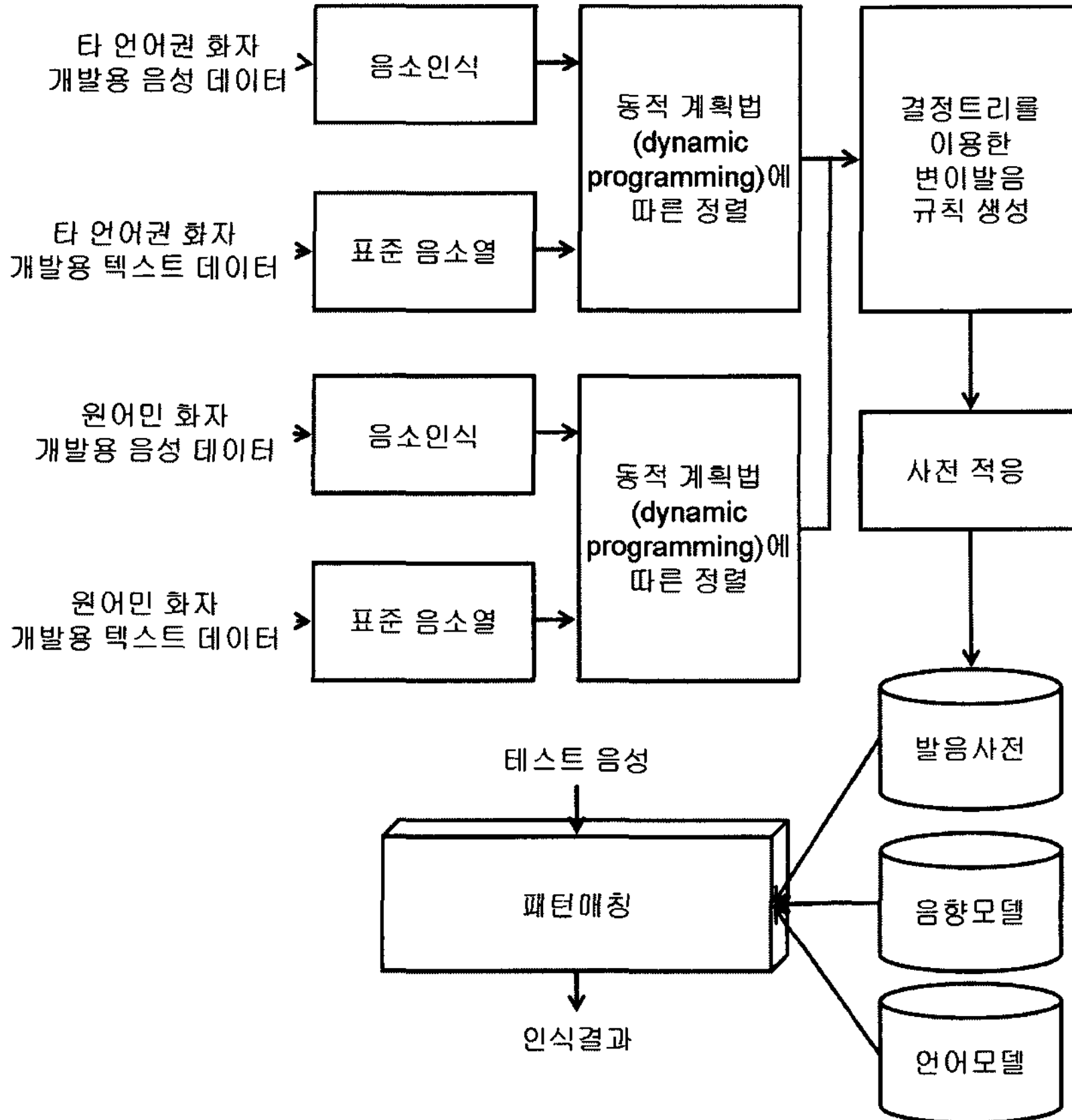
### 2.1. 음성 코퍼스

본 논문에서는 원어민 화자 음성 코퍼스와 타언어권 화자 음성 코퍼스를 사용하였다. 한국어 음성인식기를 사용하였으며, 중국인이 발화한 한국어를 타언어권 화자 음성으로 사용하였다.

대용량 어휘 한국어 낭독체 문장 음성 코퍼스인 CleanSent01[8]과 타언어권 화자가 발화한 한국어 음성 코퍼스인 F-Korean01[9]의 일부를 사용하였다. CleanSent01은 음성인식기의 음향모델 훈련과 발음 모델 적용에 사용되었다. 남자 100명, 여자 100명으로 총 200명이 20,806문장을 발화한 것으로 음향모델 훈련에 사용된 것은 170명이 발화한 17,996문장으로 30,633개의 다른 단어로 이루어져 있다. 발음모델 적용에는 음향모델 훈련에 사용되지 않은 문장 2,132문장으로 4,159개의 다른 단어로 이루어져 있다.

타언어권 화자 음성 코퍼스로 사용된 F-Korean01 중에 중국인이 발화한 한국어 음성 코퍼스를 발음모델 적용과 실험 데이터로 사용하였다. F-Korean01은 10명의 남자와 10명의 여성이 발화한 2,979문장으로 이 중 1,479문장은 발음모델 적용에 사용하고 나머지 1500문장은 실험 데이터로 사용하였다.

2.2. Baseline 음성인식기



<그림 1> 간접적 데이터 기반 발음변이 모델링 기법 적용 과정

본 논문에서 구성한 baseline 음성인식기에서는 12차 멜-켄스트럼(mel-frequency cepstral coefficient: MFCC), 로그 에너지, 이에 대한 1차, 2차 미분계수를 합쳐 39차 특징벡터를 10 ms 단위로 사용하였다. 음향모델 훈련 과정에서 켈스트럼 평균 정규화(cepstral mean normalization)와 에너지 정규화(energy normalization)를 사용하였다.

음향모델은 3개의 state를 가지는 left-to-right 모델을 사용하였고, 문맥 종속의 4개의 혼합밀도 함수 그리고 cross-word triphone 모델을 사용하였다. 음성인식기로는 HTK 버전 3.2를 사용하였다[10]. 모든 triphone 모델은 silence와 short pause를 포함한 42개의 monophone 모델을 확장하여 생성하였고, decision tree를 이용하여 state tying을 하였다[11]. 그 결과, 10,138개의 triphone 모델과 11,807개의 state들을 얻을 수 있었다.

### 3. 간접적 데이터 기반 타언어권 화자 발음변이 모델링

서론에서 언급했듯이 타언어권 화자 음성인식의 성능 향상을 위한 간접적 데이터 기반 발음변이 모델링 기법[4]을 본 장에서 설명하기로 한다. 일반적으로 타언어권 화자의 발음을 고려한 발음사전의 경우, 원어권 화자 음성인식에 영향을 주어 오히려 성능을 저하시키기도 한다. 그러나 간접적 데이터 기반 발음변이 모델링 기법을 통한 다중 발음사전 생성은 타언어권 화자 음성인식의 성능을 향상시킬 뿐 아니라 원어권 화자 음성인식에 크게 영향을 주지 않는다. <그림 1>은 간접적 데이터 기반 발음변이 모델링 기법의 과정을 설명한 것이다. 다음 5가지 단계를 통해 다중 발음사전을 생성할 수 있다.

- 1) 원어권 화자 개발 음성 데이터를 음소 인식의 결과로 얻어진 음소열을 표준 전사한 음소열과 동적 계획법에 의해 정렬시킨다.
- 2) 타언어권 화자 개발 음성 데이터 또한 같은 방법으로 정렬시킨다.
- 3) 정렬된 음소열에서 변이가 일어나는 음소 패턴을 얻는다.
- 4) 발음 변이가 일어나는 음소 패턴으로 결정트리를 이용하여 발음 변이 규칙을 생성한다[12].
- 5) 생성된 규칙을 표준 발음사전에 적용하여 새로운 다중 발음사전을 생성한다.

특히, 본 논문에서 사용한 간접적 데이터 기반 발음변이 모델링 기법에서는 음소 인식 결과를 바탕으로 한다. 일반적으로 음소 인식 성능의 성능은 좋지 않으므로, 이를 보완하기 위하여 100-best 리스트 결과를 사용한다. 또한, 3)의 과정을 통하여 획득된 음소 패턴 중, 신뢰성이 확인된 음소 패턴만을 발음 변이용 음소 패턴으로 사용한다. 이에 대한 구체적인 설명은 [4]에 기술되어 있다.

간접적 데이터 기반 발음변이 모델링 방법을 적용한 사전은 다중 발음사전이라 하며, 이를 기반으로 4장에서 설명할 최적화 방법을 적용하였다.

### 4. 제안된 발음사전 최적화 방법

다중 발음사전을 최적화하기 위해서는 추가되는 발음변이로 인한 혼잡도를 감소시키는 것이 필요하다. 이러한 혼잡도 감소 방법은 발음사전 내에서 높은 혼잡도를 가지는 발음변이들을 제거하는 것이다. 발음변이 규칙 확률과 로그 likelihood를 이용한 발음변이 규칙 확률을 이용하여 발음변이 규칙을 감소시킴으로써 발음사전의 크기를 줄였다[6]. 특히, 본 논문의 3장에서는 간접적 데이터 기반 발음변

이 모델링 과정 중 결정트리를 이용하여 발음변이 규칙을 생성하였다. 이 때, 결정트리에서 규칙 확률값에 따른 pruning 기법을 사용하여 발음변이 규칙을 감소시켰다[4]. 그러나 이러한 접근은 다중 발음사전에서 다른 단어의 발음열과의 관계를 고려하지 않은 것이다. 다시 말해, 한 단어에 여러 개의 변이 발음열이 존재하고, 이러한 발음열이 다른 단어의 발음열과 유사하거나 같은 경우가 발생하여 혼잡도를 증가시킨다. 게다가, 길이가 짧은 단어일수록 앞과 같은 경우가 발생하기 쉬울 뿐 아니라 음성인식기의 오인식되는 경우도 많다[7]. 따라서, 발음열의 길이를 고려한 혼잡도 측정 방법을 제안하였다. 다음 4.1절과 4.2절에서 제안된 혼잡도 측정 방법을 설명하고 이를 발음사전에 적용시키는 것을 보인다.

#### 4.1. 혼잡도 측정 방법

$M$ 을 다중 발음사전의 전체 집합으로 두고  $|M|$ 을 다중 발음사전  $M$ 에 포함된 단어의 수라 한다. 다중 발음사전  $M$ 에서  $i$ 번째 단어를  $W_i$ 라 하고 이 단어  $W_i$ 에 대한 변이 발음열의 수를  $N_i$ 라 한다.  $s_{i,j}$ 는  $i$ 번째 단어의 변이 발음 중  $j$ 번째 발음열을 의미한다.  $s_{i,j}$ 의 혼잡도(confusability measure: CM)를 다음 식으로 정의할 수 있다.

$$CM(s_{i,j}) = L(s_{i,j}) \times \left( \min_{\substack{1 \leq k \leq |M|, k \neq i, \\ 1 \leq l \leq N_k}} D(s_{i,j}, s_{k,l}) \times L(s_{k,l}) \right) \quad (1)$$

여기서  $D(x,y)$ 는  $x$ 와  $y$  사이의 Levenshtein distance이고 [13],  $L(x)$ 는 다중 발음사전 전체  $M$ 에서 발음열의 최대 길이  $l_{\max}$ 로 변이 발음열  $x$ 의 길이를 나누어서 정규화한 것이다. 발음열의 최대 길이  $l_{\max}$ 는 다음 식과 같다.

$$l_{\max} = \max_{1 \leq i \leq |M|, 1 \leq j \leq N_i} \#(s_{i,j}) \quad (2)$$

여기서  $\#(x)$ 는 발음열  $x$ 에 있는 음소 수를 말한다. 본 논문에서 제안한 식 (1)의 혼잡도 측정을 통한 발음사전의 최적화 방법은 다중 발음사전에서 높은 유사도를 가지는 단어의 발음열들로 인해 발생하는 음성인식기의 성능 저하를 줄일 수 있다.

##### 4.1.1. Levenshtein 거리

제안한 혼잡도 측정 방법에 사용된 Levenshtein 거리는 다음 식과 같이 정의할

수 있다[13].

$$D(x,y) = d(l_x, l_y) \quad (3)$$

여기서  $l_x = \#(x)$ 와  $l_y = \#(y)$ 는  $x$ 와  $y$ 의 음소 수를 의미하고,  $d(l_x, l_y)$ 는  $x$ 와  $y$ 의 축적된 dissimilarity를 나타내며 식 (4)와 같이 dynamic programming을 이용하여 계산된다.

$$d(n,m) = \min \left\{ \begin{array}{l} d(n-1,m) + 1, \\ d(n,m-1) + 1, \\ d(n-1,m-1) + r(x(n),y(m)) \end{array} \right\} \quad (4)$$

for  $n = 1, \dots, l_y$

식 (4)에서 각 path의 거리인  $d(n,m)$ 은  $(n,m)$ 의  $x$ - $y$  평면의 두 개의 좌표로부터 각 거리를 정의한다. 다시 말해,  $(n-1,m)$ 에서  $(n,m)$  또는  $(n,m-1)$ 에서  $(n,m)$ 으로 가는 path 거리는 deletion 또는 insertion을 의미하고,  $(n-1,m-1)$ 에서  $(n,m)$ 으로 이동하는 것은 substitution을 의미한다. Deletion과 insertion의 경우, 한 칸씩 이동할 때마다 1을 더해준다. Substitution의 경우,  $x$ 의  $n$ 번째 음소와  $y$ 의  $m$ 번째 음소가 다른 경우  $r(x(n),y(m))$ 이 식 (5)에 의해 1의 값을 가지게 된다.

$$r(x(n),y(m)) = \begin{cases} 0, & \text{if } x(n) = y(m) \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

여기서 초기값  $k = 0, 1, \dots, l_x$ 일 때  $d(k,0) = k$ 로,  $j = 0, 1, \dots, l_y$ 일 때  $d(0,j) = j$ 로 초기값을 설정한다.

#### 4.1.2. 정규화된 발음열의 음소 수

변이 발음열  $x$ 에 대한 정규화된 발음열의 음소 수를 다음 식과 같이 정의한다.

$$L(x) = \frac{l_x}{l_{\max}} \quad (6)$$

여기서  $l_x = \#(x)$ 를 의미하고 이는 앞에서 언급했듯이  $x$ 의 음소 수를 의미한다.  $l_{\max}$ 는 다중 발음사전  $M$ 에서 모든 발음열들의 음소 수 중 가장 큰 값을 의미한다. 식 (6)은 음성인식 시스템에서 발음열의 길이가 짧을수록 더 많은 오인식을 발생시키는 것을 줄이는데 기여한다.

## 4.2 혼잡도 측정을 통한 발음사전 최적화

본 절에서는 앞 절에서 정의한 혼잡도 측정을 간접적 데이터 기반 발음변이 모델링을 통한 다중 발음사전에 적용하는 방법을 설명한다. 각 다중 발음사전의 표준 발음열을 제외한 모든 변이 발음열들의 CM 값을 계산하고, 값의 순서대로 나열한다. 그런 다음, CM 값이 문턱값을 넘을 경우 남겨 두고, 그 이하의 발음열들은 혼잡도를 증가시키는 발음열로 간주하여 제거한다.

<표 1>은 “명해”라는 단어의 CM 값을 측정한 예를 나타내고 있다. 본 예에서는 이전에 간접적 데이터 기반 발음변이 모델링 방법을 제안한 논문에서 사용한 발음열 표기 방식을 사용하였다 [4]. “명해”라는 단어의 표준 발음열은 /m v N h E/이며, 세 가지 변이 발음열 /m v h E/, /m v N n E/, /m v n E/가 생성되었다. 여기서 문턱값이 0.2일 경우, /m v h E/와 /m v n E/는 혼잡도를 증가시키는 발음열로 간주되어 다중 발음사전에서 제거된다.

<표 1> 단어 “명해” (표준 발음열 /m v N h E/)의 간접적 데이터 기반 변이 발음열들의 혼잡도 측정의 예

변이 발음열	CM 값
m v h E	0.167
m v N n E	0.278
m v n E	0.167

## 5. 실험 및 결과

본 장에서는 제안한 혼잡도 감소 방법을 통한 다중 발음사전의 최적화를 음성 인식 시스템에서 성능을 측정하고, 이를 baseline 사전과 혼잡도 감소 방법을 사용하지 않은 다중 발음사전과 성능을 비교한다.

<표 2>는 3장에서 설명한 baseline 발음사전과 다중 발음사전을 사용하여 음성 인식 시스템에서 실험한 인식 결과를 보여준다. 또한 제안한 혼잡도 감소 방법을 이용한 최적화된 다중 발음사전을 여러 가지 문턱값에 따른 인식 결과를 보여준다. 또한, 음성인식 결과뿐만 아니라 인식하는 시간, 전체 CM 평균값, 사전의 크기 등을 비교하였다. 이때, 전체 CM 평균값은 발음사전에서 모든 발음열에 대한 CM 값을 더하여 계산하였다.

<표 2>에서 첫 번째 행과 두 번째 행을 비교해 보면, 간접적 데이터 기반 다중 발음사전을 사용한 음성인식 결과 단어 오인식률이 baseline 발음사전에 비해서 상대적으로 20.9% 감소하였다. 그러나 인식 시간은 실시간의 0.39배에서 0.60배로 증



가하였고, CM 값도 61.51에서 100.17로 증가하였다.

다음은 제안한 혼잡도 감소 방법을 사용하여 최적화한 다중 발음사전은 <표 2>의 3번째 행에서 보는 바와 같이 최적화하지 않은 다중 발음사전에 비해서 단어 오인식률이 더 감소하는 것을 볼 수 있다. 또한, CM 문턱값이 커짐에 따라, 다중 발음사전의 크기가 작아짐을 알 수 있다. 다시 말해, CM 문턱값이 0인 경우는 생성된 모든 변이 발음열을 사용하므로 최적화하지 않은 다중 발음사전과 동일한 반면, CM 문턱값이 커질수록 사용하지 않는 발음열이 증가하게 되므로 발음사전의 크기가 작아진다. 그리고 다중 발음사전의 CM 문턱값에 따라서 많은 실험을 한 결과, CM 문턱값이 0.06일 때 가장 좋은 성능을 보였으며, 이때 인식 결과는 baseline 발음사전에 비해 상대적으로 14.63% 성능이 향상되었고, 최적화하지 않은 다중 발음사전에 비해 상대적으로 6.25% 성능이 향상되었다. 또한, 제안한 혼잡도 감소 방법을 통한 최적화한 다중 발음사전의 인식 속도가 최적화하지 않은 다중 발음사전에 비해 상대적으로 11.67% 감소하는 것을 보여 주고 있다.

<표 2> Baseline 발음사전, 최적화 하지 않은 다중 발음사전, 제안한 혼잡도 감소 방법을 사용하여 최적화한 다중 발음사전의 성능 결과

발음사전		WER (%)	CM평균값	상대적 사전 크기	인식시간
Baseline		24.60	61.51	1.000	0.39
다중 발음사전		22.40	100.17	1.527	0.60
최적화한 다중 발음사전	문	0.03	21.60	100.15	1.518
	턱	0.06	21.00	100.07	1.501
	값	0.10	22.26	99.94	1.484

## 6. 결 론

본 논문에서 간접적 데이터 기반 발음변이 모델링 기법을 이용하여 생성한 다중 발음사전의 혼잡도를 감소시키기 위해 발음사전을 최적화 하는 방법을 제안했다. 제안된 혼잡도 감소 방법은 다중 발음사전의 변이 발음열들 중에 다른 발음열들과 혼잡도가 높은 변이 발음열을 제거함으로써 발음사전을 최적화시킬 수 있었다. 먼저, Levenshtein 거리와 발음열의 정규화된 음소 수를 이용한 혼잡도 측정 방법을 정의하고 이를 다중 발음사전에 적용하였다. 본 논문에서는 한국어 연속 음성인식기를 사용하여 제안한 방법을 평가하였다. Baseline 음성인식기를 구축하기 위하여 한국어를 원어민 화자 음성으로 사용하였고, 발음사전의 성능 평가를 위하여 타언어권 화자 음성으로 중국인이 발화한 한국어 1,479문장을 사용하였으며 baseline 음성인식기의 단어 오인식률은 24.60%이었다. 제안한 혼잡도 감소 방법을

사용하여 실험한 결과, 21.00%의 단어 오인식률을 보였다. 이는, baseline에 비해 상대적으로 14.63% 향상되었고, 혼잡도 감소 기법을 사용하지 않은 다중 발음사전에 비해 상대적으로 6.25%의 향상을 보였다. 또한, 제안한 혼잡도 감소기법을 사용하여 최적화된 다중 발음사전의 크기는 최적화하지 않은 발음사전의 크기와 비교하여 감소함으로써 인식시간의 개선을 보였다. 즉, 최적화하지 않은 발음사전을 사용할 때 인식시간이 실시간의 0.60배인 반면, 제안한 혼잡도 감소기법을 사용하여 최적화된 다중 발음사전을 사용한 경우 인식시간이 실시간의 0.53배로 감소하여 인식속도가 상대적으로 11.67% 감소함을 알 수 있었다.

### 참 고 문 헌

- [1] S. Goronzy, M. Sahakyan, W. Wokurek, "Is non-native pronunciation modeling necessary", *Proc. Eurospeech*, Vol. 1, pp. 309-312, 2001.
- [2] J. Bellegarda, "An overview of statistical language model adaptation", *Proc. ITRW on Adaptation Methods for Speech Recognition*, pp. 165-174, 2001.
- [3] I. Amdal, F. Korkmazdiy, A. C. Surendran, "Data-driven pronunciation modelling for non-native speakers using association strength between phones", *Proc. ASRU*, Vol. 1, pp. 85-90, 2000.
- [4] M. Kim, Y. R. Oh, H. K. Kim, "Non-native pronunciation variation modeling using an indirect data-driven method", *Proc. ASRU*, Vol. 1, pp. 231-236, 2007.
- [5] M. Tsai, F. Chou, L. Lee, "Improved pronunciation modeling by properly integrating better approach for baseform generation, ranking and pruning", *Proc. ISCA Workshop on Pronunciation Modeling and Lexical Access (PMLA)*, pp. 77-82, 2002.
- [6] I. Amdal, F. Korkmazkiy, A. C. Surendran, "Joint pronunciation modelling of non-native speakers using data-driven methods", *Proc. ICSLP*, pp. 622-625, 2000.
- [7] G. Hernandez-Abrego, L. Olorenshaw, R. Tato, T. Schaaf, "Dictionary refinements based on phonetic consensus and non-uniform pronunciation reduction", *Proc. ICSLP*, pp. 551-554, 2004.
- [8] 김봉완, 최대림, 김영일, 이광현, 이용주, "SiTEC의 공동 이용을 위한 음성 코퍼스 구축 현황 및 계획," *말소리*, 제46호, pp. 175-185, 2003.
- [9] 김봉완, 이용주, "음성정보기술산업지원센터의 음성 코퍼스 구축 현황 및 계획," *한국음향학회 춘계 학술대회지*, pp. 49-52, 2002.
- [10] S. Young, et al., *The HTK Book (for HTK Version3.2)*, Microsoft Corporation, Cambridge University Engineering Department, 2002.
- [11] S. Young, J. Odell, P. Woodland, "Tree-based state tying for high accuracy acoustic modeling", *Proc. ARPA Human Language Technology Workshop*, pp. 307-312, 1994.
- [12] <http://www2.cs.uregina.ca/~dbd/cs831/index.html>.
- [13] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals", *Soviet Physics Doklady*, Vol. 10, No. 8, pp. 707-710, 1966.

접수일자: 2008년 2월 19일

게재결정: 2008년 3월 24일

▶ 김민아(Mina Kim)

주소: 153-801 서울특별시 금천구 가산동 60-39 LG전자 가산 MC 연구소

소속: LG전자 MC 사업부

전화: 02) 3777-1114

E-mail: kma58@lge.com

▶ 오유리(Yoo Rhee Oh)

주소: 500-712 광주광역시 북구 오룡동 1번지 광주과학기술원

소속: 광주과학기술원(GIST) 정보통신공학과 휴먼컴퓨팅 연구실

전화: 062) 970-3121

E-mail: yroh@gist.ac.kr

▶ 김홍국(Hong Kook Kim) : 교신저자

주소: 500-712 광주광역시 북구 오룡동 1번지 광주과학기술원

소속: 광주과학기술원(GIST) 정보통신공학과 휴먼컴퓨팅 연구실

전화: 062) 970-2228

E-mail: hongkook@gist.ac.kr

▶ 이연우(Yeonwoo Lee)

주소: 전남 무안군 청계면 도림리 61번지 목포대학교

소속: 목포대학교 공과대학 정보공학부 정보통신공학전공

전화: 061) 450-2745

E-mail: ylee@mokpo.ac.kr

▶ 조성의(Sung Eui Cho)

주소: 전남 무안군 청계면 도림리 61번지 목포대학교

소속: 목포대학교 공과대학 컴퓨터교육과

전화: 061) 450-2773

E-mail: chosug@mokpo.ac.kr

▶ 이성로(Seong Ro Lee)

주소: 전남 무안군 청계면 도림리 61번지 목포대학교

소속: 목포대학교 공과대학 정보공학부 정보전자공학전공

전화: 061) 450-2436

E-mail: srlee@mokpo.ac.kr