

# 한국인을 위한 외국어 발음 교정 시스템의 개발 및 성능 평가

김무중(언어과학), 김효숙(언어과학), 김선주(언어과학),  
김병기(언어과학), 하진영(강원대), 권철홍(대전대)

## <차 례>

- |  |                          |
|--|--------------------------|
| 1. 서론                                      | 3.2. 음성 DB 수집            |
| 2. 영어 발음 교정 시스템 구성                         | 3.3. 표준 영어 음소 모델링 기법     |
| 2.1. 시스템 구성                                | 3.4. 한국인 발화데이터 음소 모델링 기법 |
| 2.2. 음소 집합                                 | 4. 실험 및 결과 분석            |
| 2.3. 발음사전                                  | 4.1. 한국인의 음소 오류 분석       |
| 2.4. 단어별 표준 발음 및 오류 발음 검출을 위한 음소 기반 인식네트워크 | 4.2. 발음 형태에 따른 인식률 분석    |
| 3. 음소 기반 음향모델링                             | 5. 결론 및 향후 연구            |
| 3.1. 영어 및 혼동 가능한 한국어 음소 기반의 음향모델링 기법       |                          |

## <Abstract>

### Performance Evaluation of English Word Pronunciation Correction System

Mu Jung Kim, Hyo Sook Kim, Sun Ju Kim, Byoung Gi Kim,  
Jin-Young Ha, Chul Hong Kwon

In this paper, we present an English pronunciation correction system for Korean speakers and show some of experimental results on it. The aim of the system is to detect mispronounced phonemes in spoken words and to give appropriate correction comments to users. There are several English pronunciation correction systems adopting speech recognition technology, however, most of them use conventional speech recognition engines. From this reason, they could not give phoneme based correction comments to users. In our system, we build two kinds of phoneme models: standard native speaker models and Korean's error models. We also design recognition network based on phonemes to detect Koreans' common mispronunciations. We get 90% detection rate in insertion/deletion/replacement of phonemes, but we cannot get high detection rate in diphthong split and accents.

\* Keywords: Speech recognition, Multi language speech DB, English pronunciation correction system for non-native speaker

## 1. 서 론

영어 학습에서 소리를 통한 학습은 학습의 효율성 및 언어 습득의 측면에서 중요한 과정 중의 하나이며, 이러한 관점에서 영어 발음 학습에 음성인식, 합성을 접목시킨 다양한 영어 학습 제품이 출시되고 있다. 그러나 기존의 영어 발음 학습 시스템은 일반 영어 음성 인식기를 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[1].

본 논문에서 제안하는 영어 발음 교정 시스템은, 한국어어를 모국어로 하는 한국인들의 한국어와, 영어의 상이한 음소체계 및 음절 구조로 인한 잘못된 영어 발음을 음소 단위로, 오류 음소를 검출 및 교정하기 위하여 음성학과 음성인식 기술을 접목시킨 시스템이다. 이 시스템은 다음 두 가지 요소로 구성되어 있다. 첫째, 표준 영어 발음을 모방하는 한국 화자들의 발음을 진단하고 평가하는 부문이 필요하다. 둘째, 한국 화자의 발음을 진단하고 평가한 결과에 따라 최적의 교정 정보를 제공하는 부문이 필요하다. 이 두 구성 요소를 결합한 시스템은 학습자와 컴퓨터 간의 학습 도구로써 상호작용 인터페이스의 기능을 제공하게 된다.

본 논문의 영어 발음 교정 시스템은 위에서 언급한 첫 번째 구성 요소인 진단과 평가 기능을 구현하는 요소 기술로써 오류 음소 검출기능 구현을 위하여 음성인식 기술을 채택하였다. 오류 음소 검출을 위하여 음성인식 기술 요소인 음소 선정, 음성 DB 구축, 그리고 음향모델 및 언어모델들이 일반 음성인식기와는 다른 구조로 설계하였다. 음소 선정은 표준 영어 음소셋과 오류 발음 유형의 음소들을 반영하기 위하여 음가는 영어 음소와 비슷하지만 음성 자질이 상이한 한국어 음소를 선정하였으며, 그리고 선정된 음소를 기반으로 발화 단어 집합을 구성하여 미국인과 한국인을 대상으로 음성 DB를 구축하였다. 구축된 음성 DB를 기반으로 음소 단위의 2가지 형태의 음향모델을 생성하였으며, 단어 내 오류 음소 검출용 음소 단위 인식네트워크를 구성하였다.

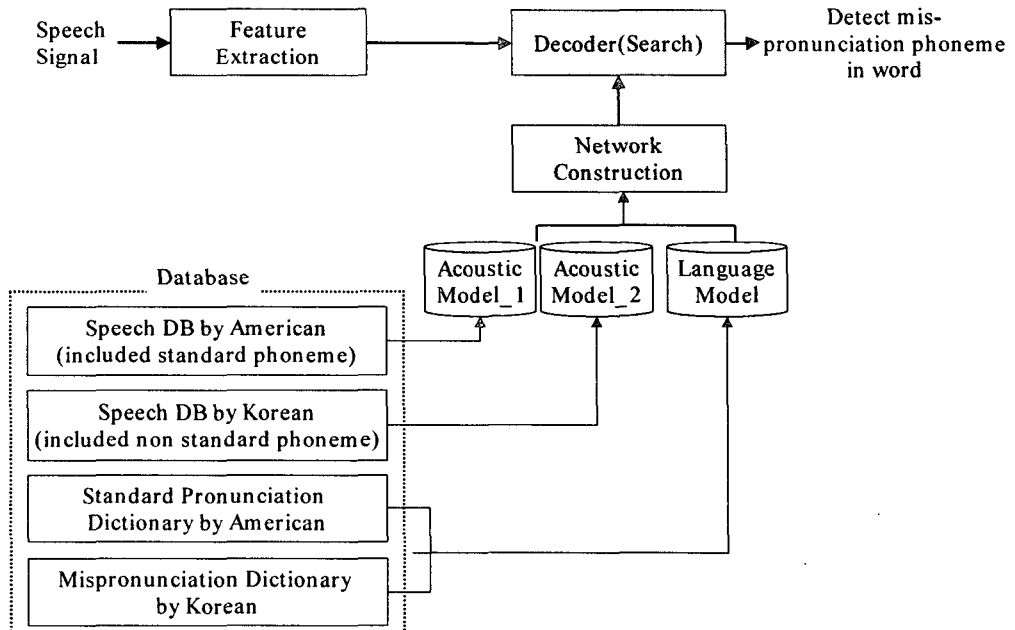
본 논문의 구성은 다음과 같다. 2장에서는 영어 발음 교정 시스템 구성 및 오류 검출을 위한 음소 기반 인식네트워크를 구축하기 전의 데이터 분석 방법과 음향모델링 구축을 위한 음성 DB와 관련 규칙 등을 살펴보고, 3장에서는 영어 발음 내 오류 음소 검출을 위한 음향모델링 기법과 음소 기반 인식네트워크 기법, 5장에서는 분석 관점에 따른 성능 분석 그리고 마지막 4장에서는 결론 및 향후 연구 과제를 논한다.

## 2. 영어 발음 교정 시스템 구성

본 장에서는 영어 발음 오류 검출 기능을 가진 영어 발음 교정 시스템의 구성 요소와 언어 네트워크와 음향모델을 생성하기 전 단계의 데이터 수집에 대한 내용을 다룬다.

### 2.1. 시스템 구성

<그림 1>은 영어 발음 교정기 시스템을 도시화한 것이다. 기존의 음성인식기와는 음향모델과 언어모델을 구성하기 전 단계인 관련 데이터베이스 구성에서 상이하며, 이를 근간으로 구성된 음향모델과 음소 단위 인식네트워크의 구성이 또한 상이하다. 인식기 내에서는 발음사전이 필요 없으며, 발음사전은 음소 단위 인식네트워크를 구성하는 자료로 사용된다[2].



<그림 1> 발음 교정기 블록 다이어그램

<그림 1>에서 2가지 형태의 음향모델을 살펴보면, 미국인이 발화한 음성 DB를 기반으로 표준 영어 음소 단위 음향모델을 생성하고, 한국인이 발화한 음성 DB를 기반으로 한국인이 발화한 표준 영어 음소들과, 표준 영어 음소와 혼동 가능한 한국어 음소 단위의 음향모델을 생성하게 된다. 음소 기반 인식네트워크의 모델인 언어모델을 살펴보면, 단어별로 표준 영어 발음사전을 기반으로 표준 영어 음소가

표기되어 있으며, 각 표준 영어 음소와 대응되는 한국인의 영어 오류 발음에 나타난 음소 셋들을 병기되어 있다.

## 2.2. 음소 집합

표준 영어 발음 사전의 구축과 음성 데이터 전사기호 사용을 위해서 음소 집합 선정이 필요하다. 시스템에서는 공개된 CMU (Carnegie Mellon University) 음소 집합을 참고하여 가공하였다. CMU 음소 집합은 자음 24개, 모음 15개(강세 미표기)의 총 39개의 음소로 이루어져 있으며, 변이음을 나타내는 음소 집합이 포함되지 않았으므로 단어 내의 변이음 규칙을 분석한 후에 나타나는 변이음 음소를 선정하였다[3, 4]. 다음은 변이음 규칙의 일부이다.

- (1) 무성과열음 /p, t, k/는 음절 초 /s/ 다음에서는 무기음화 된다.
- (2) 치조음 /d, t/는 모음 사이에 오며 두 번째 모음에 강세가 없을 때 탄설음화 되기도 한다.
- (3) 설측음 /l/은 모음 뒤나 단어 끝의 자음 앞에서 연구개음화 된다.

모음의 경우 14개의 모음에 강세와 비강세 구분을 주어 총 28개의 음소 집합을 정의하였다. 그리고, 한국어와 영어의 음운 체계, 특히 음소목록이 차이가 나기 때문에 발생하는 오류 유형인 대치, 음절 구조가 다르기 때문에 생기는 삽입 및 삭제제를 검출하기 위하여 한국어 음소들을 추가하였다. 추가된 한국어 음소 목록은 다음과 같다.

- (1) /b, d, g/ : /b, d, g/와 흔히 대치되는 한국어 음소
- (2) /s/ : /s/와 흔히 대치되는 한국어 음소
- (3) /z/ : /z, dʒ/와 흔히 대치되는 한국어 음소
- (3) /w/ : 자음군에서 흔히 삽입되는 음소

이 외에, 위의 한국어 음소 목록과는 별개로 표준 영어 발음을 잘 모방한 한국 화자의 발음이 원어민 화자의 발음과 얼마나 유사한지를 평가하기 위해 한국 화자의 발음은 표준 영어 음소에 구분기호를 넣어 이후의 분석에 이용되도록 하였다. <표 1>은 선정된 음소 집합의 구성이다.

<표 1> 영어 발음 검출시스템의 음소 집합 구성

구분	자음	변이음	모음	한국어 음소 목록	계
개수	24	16	28	6	74

### 2.3. 발음 사전

영어 발음 교정 시스템을 위한 발음사전은 두 종류이다. 첫 번째는 영어 철자에 영어 음운 규칙을 적용한 표준 영어 발음사전이다. 두 번째는 표준 영어 발음에 한국어 음운 목록과 음절 구조의 영향에 의한 변이를 포함한 한국인의 오류 발음 형태를 표현한 영어 발음사전이다. 각각의 사전을 구성하기 위해서는 표준 영어 음소 집합과 한국인의 영어 발음 음소집합이 필요하다[5, 6].

표준 영어 발음사전 구성은 북미 영어 발음을 표준으로 선택한 CMU dictionary를 2.2.절에서 선정한 음소 집합으로 변경하고, 다중 발음 단어에 대해서는 단어에 구분기호를 넣어 다중 발음단어가 포함된 표준 영어 발음 사전을 구축하였다. 한국인 영어 발음 사전은 단어별 표준 발음과 한국인의 오류 유형을 표현한 오류 발음을 한국인 영어 발음 사전으로 구축하였다. 예를 들어 단어 boy인 경우 표준 영어 발음사전은 [b oy]로 나타나져 있으며, 한국인 오류 발음사전에는 [b oy] [v oy, b o I] 등으로 오류 발음 형태가 단어별로 나타나게 된다. 이를 기반으로 구축된 표준 발음 사전과 오류 발음 사전은 오류 검출용 음소 단위 인식네트워크를 구성하는 기반 자료가 된다[7].

### 2.4. 음성 DB 수집

음성 DB를 수집하기 위하여 한국인 및 미국인 남/녀 학생들을 대상으로 한국인은 1인당 2회에 걸쳐 163단어를, 미국인은 1인당 10회에 걸쳐 2,145단어를, 동부 표준 발음을 청취 후 핸드마이크와 헤드셋을 이용해 2채널로 16KHz 샘플링률과 16비트 샘플당 비트수의 Windows PCM 포맷으로 사무실 환경에서 녹음하였다.

음향모델링의 강인성을 높이기 위하여 녹음 후 앞뒤 300msec 정도의 묵음 존재 여부 및 최대 진폭 15,000 정도의 값을 기본으로 양질의 데이터를 선별하는 작업을 하여 녹음 데이터의 약 93% 정도의 데이터를 선별하였다. <표 2>는 미국인 및 한국인에 대한 남녀 성비로의 녹음 발화자의 구성비를 나타낸다[8].

<표 2> 음성 DB 발화자 구성

구분	미 국 인			한 국 인			계
	남	녀	계	남	녀	계	
인 원	121	133	919	163	248	0	665

### 3. 오류 발음 음소 검출을 위한 음향모델링과 음소 기반 언어네트워크

3장에서는 오류 유형별 오류 검출을 위한 단어별 음소 기반 인식네트워크의 구성 방법과 미국인 표준 영어 음소와 한국인 표준 영어 음소와 혼동 가능한 한국어 음소 등의 음소 단위 음향모델링 기법을 소개한다[9].

#### 3.1. 단어별 표준 발음 및 오류 발음 검출을 위한 음소 기반 인식네트워크

영어 발음 교정시스템은 한국인이 발화한 영어 단어 내 표준 음소 발음 및 오류 음소 발음을 검출하는 기능이 있다[10]. 단어 내 오류 발음을 검출하기 위해서는 단어 내 표준 발음 음소와 오류 발음 음소가 병렬 또는 생략 가능한 것으로 연결된 단어별 음소 기반 인식네트워크가 구축되어야 한다. 다음은 오류 유형별 음소 기반 인식네트워크를 구성하는 방법이다.

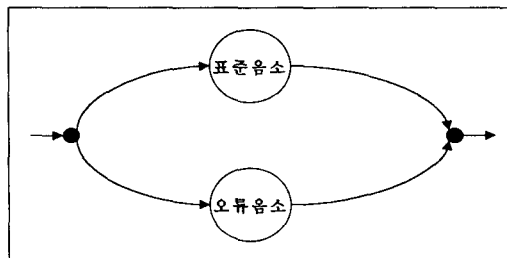
##### (1) 대치

대치 유형은 표준 영어 음소 대신 표준 영어 음소와 가장 가까운 한국어 음소로 발음하는 경우를 말한다.

영어의 순치마찰음 /f, v/는 양순파열음 /p, b/로 흔히 대치되었고, 치간무성마찰음 /θ/는 치조마찰음 /s/나 치조파열음 /t/ 등으로 대치되었다. 치간유성마찰음 /ð/는 치조 파열음 /d/ 또는 우리말 /ㄷ/ 등으로 대치되는 예가 많았다.

그리고 영어의 유성파열음 /b, d, g/가 우리말 /ㅂ, ㄷ, ㄱ/로 흔히 대치되었다. 우리말 /ㅂ, ㄷ, ㄱ/는 유성음 사이에서는 유성화가 이루어지지만 영어의 /b, d, g/와는 다르다. 무성파열음 /p, t, k/가 주로 /s/ 다음에 올 때 기식이 거의 없는 소리가 되는데 한국어 화자들은 많은 경우 기식이 있는 소리로 발음하였다.

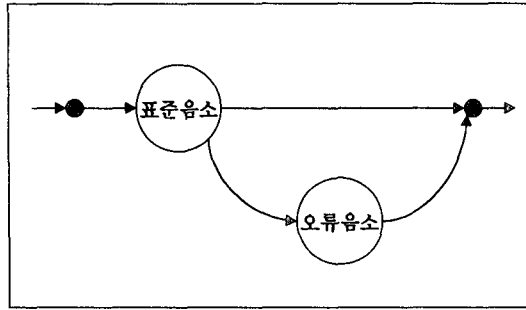
또 모음의 경우 전설모음 /i:/와 /ɪ/는 흔히 혼동되었고 후설모음 /u:/와 /ʊ/도 흔히 혼동되었다. 우리말 /에, 애/의 구분이 모호해지는 특성이 영어에도 그대로 반영되어 /æ/가 흔히 /e/로 대치되었다. 이중모음 가운데 /ou, ei/ 등은 많은 경우 /o, e/처럼 단순모음으로 대치되기도 하였다.



<그림 2> 음소 대치 오류 유형 검출을 위한 인식네트워크 구성

(2) 삽입

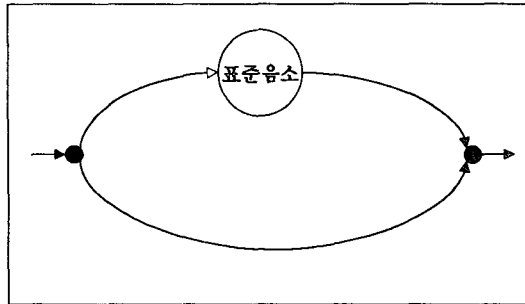
영어는 한국어와 달리 음절머리 또는 음절꼬리에 자음군을 허용하고 있다. 'strike'라는 단어는 한국인이 발음하면 흔히 [스트라이크]가 된다. 음절머리에 /str/ 처럼 세 개의 자음이 연이어 오는 현상이 한국어에 없기 때문에 [으] 모음이 흔히 삽입되는 것이다. 그리고 'sports'와 같이 음절꼬리에 오는 자음군에도 [으]를 삽입하여 [스포츠]처럼 흔히 발음한다.



<그림 3> 음소 삽입 오류 유형 검출을 위한 인식네트워크 구성

(3) 삭제

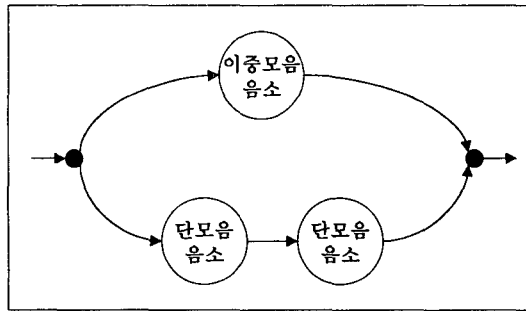
단어 끝에서 자음 [r]이 삭제되는 경우가 있으며, /q/다음에 오는 반자음 [w]이 삭제되는 경우이다.



<그림 4> 음소 삭제 오류 유형 검출을 위한 인식네트워크

(4) 이중모음 분리

한국어에는 존재하지 않는 /oy ai au ou/ 등의 음소 발음시 한국인은 이를 분리하여 발음하는 경향이 있다. 이를 검출하기 위한 음소 기반 인식네트워크가 <그림 5>에 나타나있다.



<그림 5> 음소 삭제 오류 유형 검출을 위한 인식네트워크

### 3.2. 영어 및 혼동 가능한 한국어 음소 기반의 음향모델링 기법

음향모델은 선정된 음소를 기준으로 한국어 화자에 의해 발화된 데이터와 미국인 화자에 의해 발화된 데이터를 구분하여 각각 음소 단위 HMM 모델을 생성 후 통합시키는 방식을 사용하였다. 미국인과 한국어인 음성데이터에 대해서 pre-emphasis와, 10msec마다 25msec 크기의 Hamming window를 적용하여 특징벡터를 추출하였다. 특징 벡터는 사람의 청각기관의 특징을 고려한 MFCC (Mel frequency Cepstral Coefficients)를 사용하여 12차 MFCC와 log energy 그리고 각각의 delta 및 delta-delta를 구한 39차 특징 벡터를 사용하였다. HMM 모델은 left-to-right 형식의 3개의 상태를 기반으로 하였다. 한국어 음성 데이터에 대한 음소 집합과 외국인 음성 데이터에 대한 음소 집합이 상이하기 때문에, 다른 방식으로 2 종류의 음향 모델을 생성하였다[11].

#### 3.2.1. 표준 영어 음소 모델링 기법

미국인 발음을 반영한 음향모델은 전체 음성 데이터에 대한 global mean과 global variance로 프로토 타입을 구성 후, Baum-Welch re-estimation 알고리즘을 이용하여 표준 영어 음소 집합을 구성하고 있는 68개 음소 단위 1차 음향모델을 생성하였다. 생성된 음향모델을 기반으로 음소 경계 시간 정보를 추출하여, 시간보정 과정을 거친 후, K-means 알고리즘을 통해 음소 단위 초기 음향모델을 생성한 후, Baum-Welch re-estimation 알고리즘을 통해 2차 최종 모델을 선정하였다.

#### 3.2.2. 한국어인 발화 데이터 음소 모델링 기법

한국인이 발음한 음성 데이터는 오류 발음 유형이 포함되어 있기 때문에 혼동 가능한 한국어 음소를 반영한 모델을 만들기 위해서는 3단계를 거쳐 최종 모델링을 하였다.



첫 번째 단계는 생성된 영어 음소 단위 음향모형을 기반으로 자동 세그멘테이션 작업 후, 한국인 영어 음성 DB의 30%의 음성 데이터에 대해 음성학 전문가들의 수작업을 통해 오류 유형 분석 및 전사 그리고 시간 정보 수정 등의 매뉴얼 세그멘테이션 작업을 하였다. 두 번째 단계는 수작업된 30%의 음성데이터에 대해 K-means 알고리즘을 이용하여 74개 음소 단위 HMM의 초기 음향모형을 생성 후, Baum-Welch re-estimation 알고리즘을 이용하여 2차 음향모형을 생성한다. 세 번째 단계는 생성된 모델을 이용하여 나머지 70%의 음성 데이터의 자동 세그멘테이션 후, 30% 매뉴얼 세그멘테이션 작업된 음성 데이터들과 통합하여 K-means 알고리즘을 이용하여 초기 음향모형을 생성한 후, Baum-Welch re-estimation 알고리즘을 이용하여 최종 모델을 생성한다.

미국 원어민 음성 데이터와 한국인 발화 음성 데이터로 생성된 데이터를 이용한 인식 실험에서 적절한 mixture 개수를 찾은 결과, 미국 원어민의 데이터에 대해서는 mixture 7로 한국인의 데이터에 대해서는 mixture 10으로 최종 모델을 생성하였다[11, 12].

#### 4. 실험 및 결과 분석

영어 발음 교정 시스템의 성능 분석을 위하여, 음성 DB를 음성학 전문가들의 청취 후 판단 과정을 통하여 한국인의 발음 오류 분석을 하였으며, 음소 대치, 삽입, 삭제, 이중모음 분리 그리고 강세 오류 등의 관점에서 성능을 분석하였다.

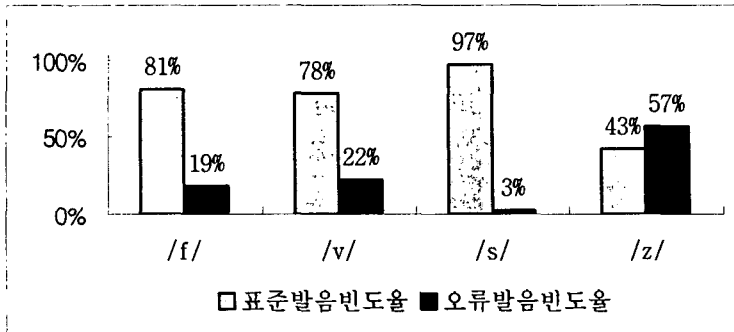
##### 4.1. 한국인의 음소 오류 분석

한국인들의 오류 유형을 분석하기 위하여 한국인 음성 DB 중 120명분의 자료 가운데 녹음 상태가 좋지 않은 것 등을 제외하고 최종 선택한 단어 개수는 20,160개(남 8,400 단어/ 여 11,760 단어)이다.

20,160개 단어를 2명의 음성학 전문가가 수작업으로 분절하였고 그 가운데 17,460개의 음소는 두 명의 작업 결과를 교차시켜 최종 확인한 것이다. 다음은 교차 확인을 마친 17,460개의 음소를 자음과 모음으로 나누어 정오 빈도 및 오류 유형별 빈도를 분석한 것이다. <그림 6, 7, 8, 9>는 한국 화자의 영어 단어 발음 분석하여 표준 영어의 음소와 비교한 결과의 일부이다.

##### (1) 대치

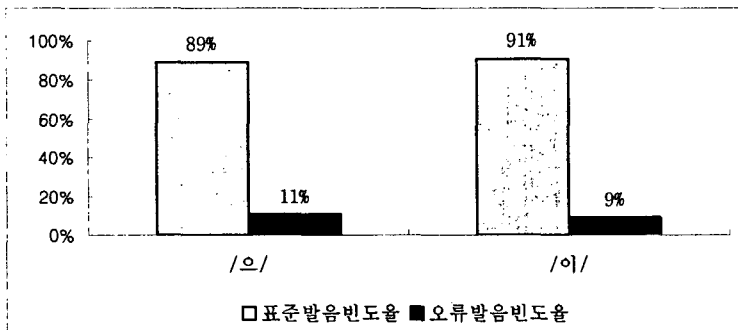
<그림 6>은 4개의 표준 음소에 대하여, 한국인 음성 DB를 대상으로 오류 유형이 나타난 빈도를 나타낸 것이다.



<그림 6> 음소 대치 오류 유형 빈도

(2) 삽입

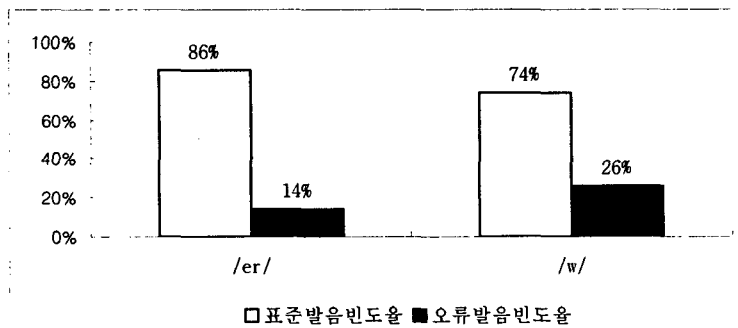
<그림 7>은 한국인 음성 DB를 대상으로 /으,이/의 삽입 빈도를 나타낸 것이다.



<그림 7> /으//이/음소 삽입 오류 유형 빈도

(3) 삭제

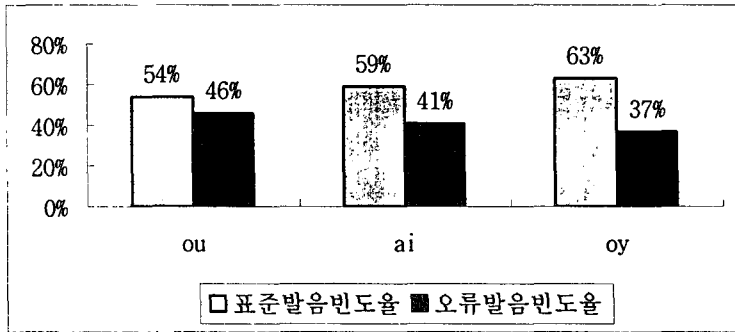
<그림 8>은 한국인 음성 DB를 대상으로 /r,w/ 음소의 삭제 빈도를 나타낸 것이다.



<그림 8> /r, w/ 음소 삭제 오류 유형 빈도

(4) 이중모음 분리

<그림 9>는 한국인 음성 DB를 대상으로 이중모음 중 /ou, ai, oy/에 대한 이중모음 분리에 대한 표준 발음 빈도 및 오류 발음 빈도를 나타낸 것이다.



<그림 9> /ou, ai, oy/에 대한 이중모음 분리 오류 유형 빈도

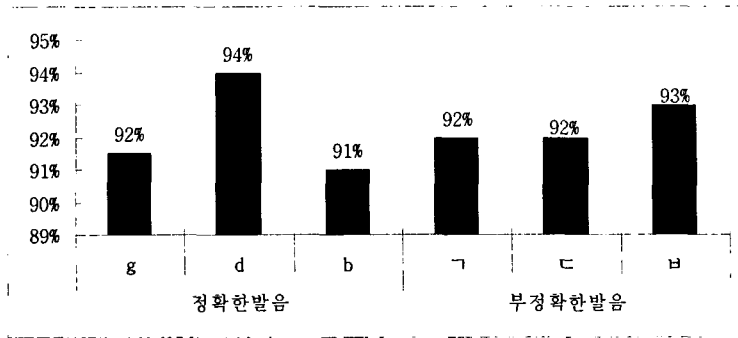
4.2. 발음 형태에 따른 인식을 분석

본 논문에서는, 분석하는 관점을 표준 음소에 대한 오류 음소 대치, 삽입, 삭제, 이중모음 분리, 강세의 5개 오류 유형에 대하여 정확한 미국 원어민 발음과 오류 유형이 포함된 한국인 발음을 입력 데이터로 구성하여 인식결과를 구하고, 이를 실험음성학 전문가들의 검토를 통하여 정검출과 오검출율로 성능을 테스트 하였다.

음소의 삽입, 삭제, 대치, 이중모음 분리, 강세 등의 오류 유형별로 대표되는 단어 10개를 선정한 후, 각 단어별로 1차로 실험음성학 전문가들이 선정한 정확한 발음 10개, 부정확한 발음 10개의 데이터를 준비하여, 오류 유형별 정확한 발음 100개, 부정확한 발음 100개의 테스트 데이터를 구성하였다.

(1) 대치

대치되는 오류 유형의 성능을 테스트하기 위하여 /g, d, b/로 시작하는 단어인 bus, dish, gold 등의 단어를 선정 후, 정확한 발음 입력 시는 /g, d, b/ 등의 음소가 인식이 되어 정검출이 되는 것이며, 부정확한 발음 입력 시는 /ㄱ, ㄷ, ㅂ/가 출력이 되어 부정확한 발음의 정확한 검출(정검출)이 되는 것이다.

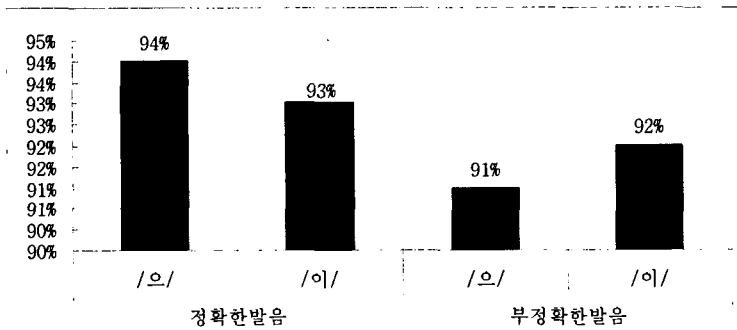


<그림 10> 음소 대치 오류 인식 시 검출률

<그림 10>은 음소 대치 오류 검출에 대한 성능 분석 결과이다. 미국인이 /g, d, b/로 시작하는 단어를 입력하였을 경우, 정확하게 /g, d, b/ 음소로 검출한 확률이 각각 92%, 94%, 91%이며, 한국인 발화 데이터 중 /g, d, b/로 발음을 해야 하는데, /ㄱ, ㄷ, ㅂ/로 발음된 오류 유형 음소를 포함한 데이터를 입력하였을 때, /ㄱ, ㄷ, ㅂ/로 검출한 확률이 각각 92%, 92%, 93%이다.

## (2) 삽입

삽입되는 오류 유형의 성능을 테스트하기 위하여 삽입 현상이 발생할 수 있는 strike, church 등의 단어를 선정하였다. 정확한 발음 인식 시에는 [으][이]의 음소가 검출되지 않아야 하며, 부정확한 발음 인식 시에는 [으] 또는 [이]가 출력되어 정 검출이 되는 것이다.

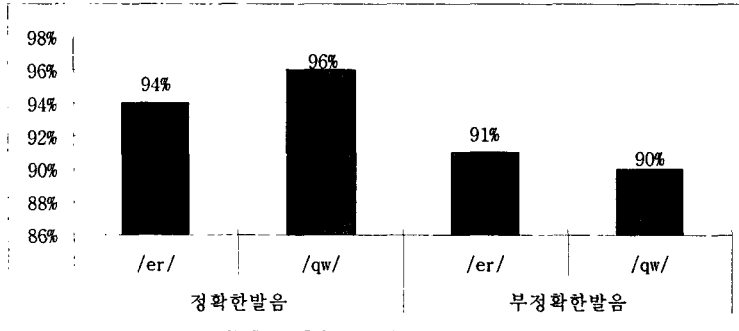


<그림 11> /으, 이/ 삽입 오류 인식 시 검출률

<그림 11>은 음소 삽입 오류 검출의 성능 분석 결과이다. 미국인이 발음한 /으, 이/ 가 삽입이 안된 표준 단어를 입력하였을 경우, /으, 이/ 음소 미검출률이 각각 94%, 93%이고, /으, 이/ 음소가 삽입된 오류 유형 포함한 한국인 데이터를 입력하였을 경우, /으, 이/의 삽입 오류 검출률이 각각 91%, 92%이다.

(3) 삭제

삭제 오류 유형의 성능을 테스트하기 위하여 teacher, quiz, wood 등의 단어를 선정하였다. 각각의 단어를 발음할 때에 정확한 발음은 /r, w/ 음소가 인식되어야 하며, 부정확한 발음 입력에는 /r, w/ 음소가 인식되지 않아야 한다.

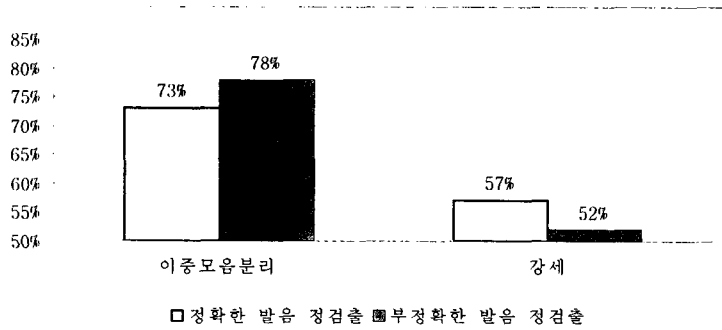


<그림 12> /r, w/에 대한 음소 삭제 오류 검출률

<그림 12>는 /er/에서 /r/ 음소삭제, /qw/에서 /w/ 음소삭제 오류 검출의 성능 분석 결과이다. 미국인의 정확한 발음을 입력하였을 경우 미삭제로 검출된 데이터가 각각 94%, 96%를 나타내며, 한국인의 /r, w/ 음소가 약하게 발음됐거나 삭제된 경우 삭제 오류 유형 검출률이 각각 91%, 90%를 나타내었다.

(4) 이중음소 분리 및 강세

이중모음 분리의 오류 유형 성능을 테스트하기 위하여 fine, town, toy, slow 등의 10개 단어를 선정하였다. 부정확하게 발음한 경우에는 이중모음이 분리되어 인식된다. 강세는 음소 집합에 제 1강세와 미강세 모음을 선정하여 모델을 생성하였으므로 강세 모음대신 미강세 모음이 인식될 경우 잘못된 검출(오검출)로 본다.



<그림 13> 이중모음 분리 및 강세 오류 검출률

<그림 13>은 이중모음 분리 및 강세 오류 검출에 대한 시스템 성능 분석 결과이다. 원어민의 이중모음 분리가 일어나지 않은 표준 발음을 입력 시 정확하게 이중모음 발음으로 검출한 확률이 73%이며, 한국인의 이중모음을 분리하여 [단모음 + 단모음]으로 발음하였을 때 [단모음 + 단모음] 분리로 검출한 확률이 78%이다. 강세에 대한 성능 부분은 미국인의 정확한 강세가 있는 데이터를 입력하였을 때 정확한 강세 데이터로 검출한 확률이 57%이며, 한국인의 부정확한 데이터를 입력한 경우 52%가 부정확한 강세 입력으로 검출하였다.

## 5. 결론 및 향후 연구

본 논문에서 제안한 영어 음소 및 혼동 가능한 한국어 음소 단위 음향모델과 오류 발음 분석을 통한 음소 단위 인식네트워크를 기반으로 한 한국인의 영어 오류 발음 검출 시스템에서 오류 형태인 대치, 삽입, 삭제 형태는 인식을 통한 검출률이 90% 이상으로 높았으나, 모음에 대한 이중모음 분리 및 강세 오류의 인식을 통한 검출률은 75%이하로 낮았다. 이는 한국인 음성 DB 구축 시 모음에 대한 매뉴얼 세그멘테이션 작업의 부족과 훈련 데이터에 대한 에너지 불균형에 문제로 추측된다.

이와 같은 낮은 검출률을 보완하기 위하여, 이중모음 분리는 해당 모음에 대한 duration 검출과 표준 음소의 duration의 비교 방식을 통한 방법을 연구 중이며, 강세는 모음에 대한 피치 및 에너지를 통한 새로운 접근 방식이 상기 결과를 보완하는 방법으로 연구 진행 중이다. 추후 리듬적인 요소 및 문장 내 발음 교정 기술의 연구는 영어 발음 학습 적용에 있어서 단어 단위가 아닌 구 및 문장 단위의 발음 학습 기법의 연구가 되어야 한다.

## 참 고 문 헌

- [1] J. Pernstein, "Automatic Grading of English Spoken by Japanese Students", *SRI International Internal Reports Project 2417*, 1992.
- [2] L. Neumeyer, H. Franco, et al., "Automatic text-independent pronunciation scoring for foreign language student speech", *Proc. of ICSLP 96*, pp.1457-1460, 1996.
- [3] P. Ladefoged, *A course in phonetics (4th edition)*, Harcourt College Publishers, 2001.
- [4] <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- [5] Bonaventura, Patrizia et al., "Phonetic annotation of a non-native speech corpus", in *Proceedings Workshop Integrating Speech Technology in the (Language) Learning and Assistive Interface*, pp.10-17, Dundee (UK), InStil 2000.
- [6] R. Sproat, "Pmttools: A pronunciation modeling toolkit," in *Proceedings of the Fourth ISCA Tutorial and Research Workshop on Speech Synthesis*, pp.153-159, Scotland: Blair Atholl, 2001.
- [7] C. Nick, J-P. Martens, "On the use of pronunciation rules for improved word recognition", in *Proceedings Eurospeech-95*, pp.1747-1750, 1995.
- [8] J. Bernsten, M. Cohen et al., "Automatic evaluation and training in English pronunciation", *Proc. of ICSLP 90*, pp.1456-1459, Kobe, Japan, 1990.
- [9] H. Niemann, W. Eckert, F. Gallwitz, "Combining stochastic and linguistic language models for recognition of spontaneous speech", *Prod. of ICASSP 96*, pp.423 - 426, Atlanta, USA, 1996.
- [10] Y. J. Chung, C. K. Un, "Use of different numbers of mixtures in continuous density hidden Markov Models", *IEE Letters*, Vol. 29, No. 9, pp.824-825, 1993.
- [11] 김무중, 권철홍, "HMM 및 보정 알고리즘을 이용한 자동 음성 분할 시스템", *음성과학*, 7권, 4호, 한국음성과학회, pp.256-274, 2002.

접수일자: 2003년 06월 01일

게재결정: 2003년 06월 12일

▶ 김무중(Mu Jung Kim)

주소: 151-716 서울시 관악구 봉천4동 882-5 2F (주)언어과학

소속: (주)언어과학 부설 언어공학연구소

전화: 02) 887-8125 (내선 310)

FAX: 02) 887-8127

E-mail: donaldos@eoneo.co.kr

## ▶ 김효숙(Hyo Sook Kim)

주소: 151-716 서울시 관악구 봉천4동 882-5 2F (주)언어과학

소속: (주)언어과학 부설 언어공학연구소

전화: 02) 887-8125 (내선 310)

FAX: 02) 887-8127

E-mail: hoyssook@eoneo.co.kr

## ▶ 김선주(Sun Ju Kim)

주소: 151-716 서울시 관악구 봉천4동 882-5 2F (주)언어과학

소속: (주)언어과학 부설 언어공학연구소

전화: 02) 887-8125 (내선 310)

FAX: 02) 887-8127

E-mail: sunju@eoneo.co.kr

## ▶ 김병기(Byung Gi Kim)

주소: 151-716 서울시 관악구 봉천4동 882-5 2F (주)언어과학

소속: (주)언어과학 부설 언어공학연구소

전화: 02) 887-8125 (내선 310)

FAX: 02) 887-8127

E-mail: killm2@eoneo.co.kr

## ▶ 하진영(Jin-Young Ha)

주소: 200-701 강원도 춘천시 효자2동 192-1 강원대학교

소속: 강원대학교 전기전자정보통신공학부

전화: 033) 250-6386

E-mail: jyha@kangwon.ac.kr

## ▶ 권철홍(Chul-Hong Kwon)

주소: 300-716 대전시 동구 용운동 96-3 대전대학교

소속: 대전대학교 정보통신인터넷공학부

전화: 042) 280-2555

E-mail: chkwon@dju.ac.kr